Exploiting Image Sensor Data in Biometric Systems and Mobile Applications

by
Luca Debiasi

Cumulative dissertation submitted to the
Faculty of Natural Sciences, University of Salzburg
in partial fulfillment of the requirements
for the Doctoral Degree.

Thesis Supervisor
Univ.-Prof. Mag. Dr. Andreas Uhl

Department of Computer Sciences
University of Salzburg
Jakob Haringer Str. 2
5020 Salzburg, AUSTRIA

Salzburg, March 2020

Abstract

In modern society, the exchange of information plays a vital role for many kinds of interactions. Many of those interactions, being human or with a computer system, require authenticating ones' identity: withdrawing money from an ATM, communicating via social media, logging into a computer system, passing a border to a foreign country, signing documents – to mention a few examples.

Biometric systems offer an alternative to classical authentication methods such as passwords, PIN codes or smart-cards in this context. These systems enable authenticating oneself by using one or multiple biometric traits, i.e. physical, chemical or behavioural characteristics. Biometric authentication systems have already established themselves in everyday life, in particular since their implementation in mobile devices such as smartphones. Besides biometric authentication, mobile devices offer a much wider and ever-increasing range of use cases and applications, which is mainly driven by the rapid technological advancements in processing power paired with a decrease in power consumption as well as miniaturisation of diverse types of sensors.

This thesis covers my research with respect to exploiting image sensor data in biometric systems and mobile applications. Several forensic methods and techniques have been adopted and adapted to biometric systems. The main focus laid on investigating whether individual biometric sensors can be differentiated and thus allowing for authentication of the source sensor used to acquire an image processed in a system. Authenticating the source sensor, i.e. the sensor that captured an image, is expected to improve the security by preventing specific attacks as well as yield recognition performance improvements by enabling sensor specific image enhancements in biometric systems. The sensor's differentiability was investigated with several device identification and clustering techniques based on the *photo response non-uniformity* (PRNU) in conjunction with different enhancements for improving the extracted PRNU's quality. In order to evaluate the experimental results, we furthermore identified suitable and reliable metrics using data with known ground truth. Apart from that, an alternative to the classical PRNU-based identification of source sensors was proposed.

Further work focused on the detection of face morphing attacks, a recently presented attack on face recognition systems exploiting their generalisability. We proposed a PRNU-based morphing detection scheme that analyses spectral and spatial characteristics of an image's PRNU to detect variations introduced by the morphing process. The system was evaluated using a variety of morphing algorithms, including landmark and neural network based ones, image processing to conceal the morphing procedure and images acquired with a large number of different cameras. In addition, the detection performance in a print-scan scenario was analysed.

The mobile applications covered in this thesis consist of an evaluation of face recognition systems for smartphones, where usability and security against presentation attacks has been analysed, the design of a near infrared (NIR) illumination prototype for smartphones that enables to capture vascular patterns as well as drug counterfeit detection from its packaging using images acquired with a smartphone.

Abstract (German)

In unserer modernen Gesellschaft spielt der Informationsaustausch eine essentielle Rolle für vielerlei Arten von Interaktionen. Viele dieser Interaktionen, ob nun zwischenmenschlich oder mit einem Computersystem, erfordern eine Authentifizierung der eigenen Identität: Bargeldabhebungen an einem Geldautomaten, Kommunikation mittels sozialer Medien, Einloggen in ein Computersystem, Überschreiten einer Grenze zu einem anderen Land, Unterschreiben von Dokumenten – um einige Beispiele zu nennen.

Biometrische Systeme bieten sich in diesem Kontext als Alternative zu klassischen Verfahren wie Passwörtern, PINs und Smart-cards an. Biometrische Systeme ermöglichen die Authentifizierung mittels eines oder auch mehreren biometrischen Merkmalen, also physischer, chemischer oder verhaltensspezifischer Eigenschaften. Der Durchbruch biometrischer Systeme hat, wenn man unseren Alltag betrachtet, bereits stattgefunden, insbesondere seit deren Einführung in mobilen Geräten wie Smartphones. Neben biometrischer Authentifizierung bieten Smartphones jedoch eine weit größere und stetig wachsende Zahl an Anwendungsszenarien.

Diese Dissertation gibt einen Einblick in meine Forschungsarbeit hinsichtlich des Ausnutzens von Bildsensor-Daten in biometrischen Systemen und mobilen Anwendungen. Hierfür wurden verschiedene forensische Methoden und Verfahren an die Verwendung in biometrischen Systemen angewendet und angepasst. Dabei wurde ermittelt, ob und inwiefern einzelne biometrische Sensoren unterschieden werden können. Dadurch wird eine Authentifizierung des Ursprungs-Sensors eines Bildes, welches in einem System verarbeitet wird, ermöglicht. Durch die Authentifizierung des Ursprungs-Sensors, d.h. des Sensors mit dem ein Bild aufgenommen wurde, lässt sich eine Verbesserung der Sicherheit sowie eine Steigerung der Erkennungsleistung in einem biometrischen System erzielen. Die Unterscheidbarkeit der Sensoren wurde mittels unterschiedlicher Methoden zur Erkennung sowie zum Clustering von Ursprungs-Sensoren ermittelt. Um die experimentellen Ergebnisse auswerten zu können, wurden geeignete und zuverlässige Metriken unter Verwendung von Daten mit bekannten Ursprungs-Sensoren eruiert. Darüber hinaus wurde auch eine alternative Methode zur Erkennung des Ursprungs-Sensor vorgeschlagen.

Ein weiterer Bereich meiner Forschung befasste sich mit der Erkennung von Gesichtsmorping-Angriffen, einer erst kürzlich vorgestellten Art von Angriffen auf Gesichtserkennungssystemen. Hierzu wurde ein Morphing-Erkennungsverfahren basierend auf der *photo response non-uniformity* (PRNU) entwickelt, das durch den Morphing-Prozess verursachte Veränderungen an der PRNU im Spektral- und Bildbereich analysiert. Das vorgeschlagene Verfahren wurde unter Einbeziehung mehrerer Morphing-Algorithmen (basierend unter anderem auf markanten Gesichtspunkten sowie neuronalen Netzen), verschiedener Bildverbesserungen (um das Morphing zu verschleiern) und Bildern einer großen Anzahl an unterschiedlichen Kameras ausgiebig evaluiert. Zusätzlich wurde die Erkennungsleistung in einem Szenario analysiert, bei dem die gemorphten Bilder zuerst ausgedruckt und dann wieder eingescannt wurden.

Bei den mobilen Anwendungen, die in dieser Dissertation behandelt werden, haben wir die Benutzbarkeit und Sicherheit von Gesichtserkennungssystemen für Smartphones ausgewertet, einen Beleuchtungsaufsatz-Prototypen für Smartphones zur Aufnahme von Venenmustern im nah-infrarot (NIR) Bereich entwickelt sowie eine Erkennung von gefälschten Medikamenten entwickelt, die auf Smartphone-Fotos der Medikamentenverpackung basiert.

Acknowledgments

WHEN A USER TAKES A PHOTO, THE APP SHOULD CHECK WHETHER THEY'RE IN A NATIONAL PARK...
SURE, EASY GIS LOOKUP. GIMME A FEW HOURS.
...AND CHECK WHETHER THE PHOTO IS OF A BIRD.
I'LL NEED A RESEARCH TEAM AND FIVE YEARS.
IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

Comic by Randall Munroe (`xkcd.com`)

The path towards this thesis was not always an easy one, but I am very grateful for the opportunity of working in such an interesting and future-oriented field.

I wish to express my deepest gratitude to my advisor Andreas Uhl, who supported and guided me through my Bachelor's, Master's and now my PhD degree. I wish to thank all the people of my research group, the Multimedia Signal Processing and Security Lab (Wave-Lab) at the University of Salzburg. In particular, I would like to thank my colleagues and co-authors Heinz Hofbauer, Christof Kauba, Simon Kirchgasser, Bernhard Prommegger and Christian Rathgeb (in alphabetical order) for the many inspiring discussions and fruitful collaboration.

This thesis would not have been possible without funding from the Austrian Science Fund FWF as part of the Biometric Sensor Forensics project, the European Union's Horizon 2020 projects PROTECT, IDENTITY and COST Action IC 1106 as well as support of Veridos GmbH and eMundo GmbH.

Last but not least, I would like to recognise the invaluable assistance that my family and friends provided during my studies. My deepest appreciation goes to my partner Lisa for her continuous support and encouragement.

Salzburg, February 2020 *Luca Debiasi*

Contents

1. Introduction

Authenticating oneself has become part of every-day life. Some examples for such authentication processes include signing documents, logging in into a computer, withdrawing money from an automated teller machine (ATM) or unlocking a smartphone. Classic authentication techniques such as signatures or passwords and PIN codes are traditionally used in these kind of scenarios, in some cases also token based systems like Smart Cards are utilised. However, these classical authentication methods pose some disadvantages: passwords can get disclosed or forgotten, tokens can be lost or stolen and signatures can be forged. Genuine users are therefore not able to authenticate themselves any longer or even worse, an impostor is able to be authenticated as a genuine user.

This thesis focuses on exploiting image sensor data in biometric systems and mobile applications, i.e. analysing the data acquired in biometric systems and mobile applications and using it for a different purposes. The exploited data might be used to enhance a system's security or detect misuse of the system. The use of mobile devices, in particular smartphones, is increasing and so does their use for various applications such as biometric authentication in mobile environments. However, these mobile systems can not only be used for authenticating a user, but enable ample other applications such as authenticating products with a smartphone as done by Authentic Vision[1] or an augmented reality experience where real world objects are enhanced by additional computer generated information.

In the following sections, the nomenclature and procedures used in biometric systems, for attacks on biometric systems and in digital image forensics as well as how they are intertwined will be explained in more detail.

1.1. Biometrics and Biometric Systems

A biometric recognition system relies on behavioural or biological characteristics of an individual, e.g. iris, face, fingerprints or the voice, and utilises these biometric traits to enhance the security and convenience for the user. Because of the individual nature of the biometric traits, which cannot be lost or be forgotten, the user is no longer required to remember complex passwords nor carry around Smart Cards. Such systems are deployed as authentication systems in industrial settings and high-security areas such as laboratories, banks and border control for several years, but their use is becoming more and more widespread in everyday life as well, i.e. fingerprint recognition in the case of smartphones like Apple's "TouchID"[2] or Microsoft's "Windows Hello"[3] for biometric authentication using fingerprint and face recognition in Windows 10. Another important application is the Indian UIDAI program "Aadhaar"[4] issuing a unique identification number to each Indian resident and using different biometric modalities to identify and distinguish the enrolled subjects.

A typical biometric system consists of three main components: a biometric sensor to capture the raw biometric data, a feature extractor that converts the raw data to a feature based representation and a matcher which compares two sets of features and yields a score value

[1] https://www.authenticvision.com

[2] https://support.apple.com/en-us/HT201371

[3] https://support.microsoft.com/en-us/help/17215/windows-10-what-is-hello

[4] https://uidai.gov.in

corresponding to the similarity or dissimilarity of the the feature sets. An individual can be registered in the system once and then be authenticated repeatedly. Biometric sensors deployed in practical applications often contain a digital image sensor to acquire images of the biometric traits. They are often adapted for the acquisition of a specific trait to improve the quality of the acquired images. For example iris sensors are mostly supported by a near infra-red (NIR) light source to improve the iris recognition results [12].

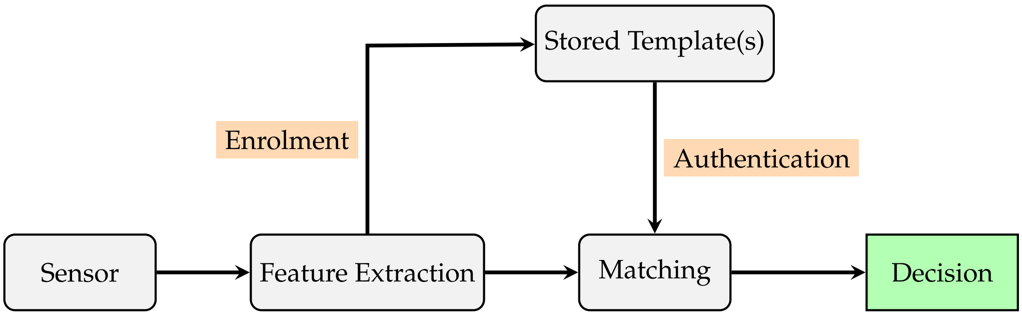


Figure 1.1.: Exemplary generic biometric system.

As shown in figure 1.1, first of all the biometric trait (e.g. the finger) is presented to the sensor (also called biometric scanner). The sensor captures a sample of the biometric trait and creates a digital representation (usually an image or video). This biometric sample is then pre-processed in order to enhance its quality. After the pre-processing, features describing the specific kind of biometric data are extracted. These features are stored in a defined format, called biometric template. Biometric recognition systems work in two stages. During enrolment, the user's biometric trait is captured, a biometric template is generated and stored in a database. During the authentication phase (either verification or identification), a sample of the user's biometric trait is captured and a new template is created. This template (also called probe template) is then compared against the templates stored in the database (also called gallery templates), resulting in a comparison score. Based on a threshold, the decision module outputs the final decision if the user is a genuine one (successfully authenticated) or an impostor one (authentication failed). For the verification scenario, a user claims an identity and his probe template is compared against the stored gallery templates of the claimed identity. Identification, as second type of authentication, compares the probe template of the user against all gallery templates stored in the database to establish the unknown identity of the user based on the highest comparison score with one of the stored templates.

1.2. Attacks on Biometric Systems

Despite their advantages over traditional authentication systems, biometric recognition systems are far from being perfect in terms of accuracy, reliability, security and thus, usability. Ratha *et al.* [56] identified eight vectors in a generic biometric system where attacks may occur as illustrated in Figure 1.2. In this work the focus lies on attacks on the first two attack vectors of a biometric system, i.e. presentation attacks (1) and insertion attacks (2), as well as the stored templates (6).

A biometric trait can be forged or spoofed, mimicking a genuine user in order to be successfully authenticated in a biometric system. These types of attacks, exploiting attack vector 1, are denoted as "presentation attacks" (PA) and are defined in the ISO/IEC 30107-1:2016

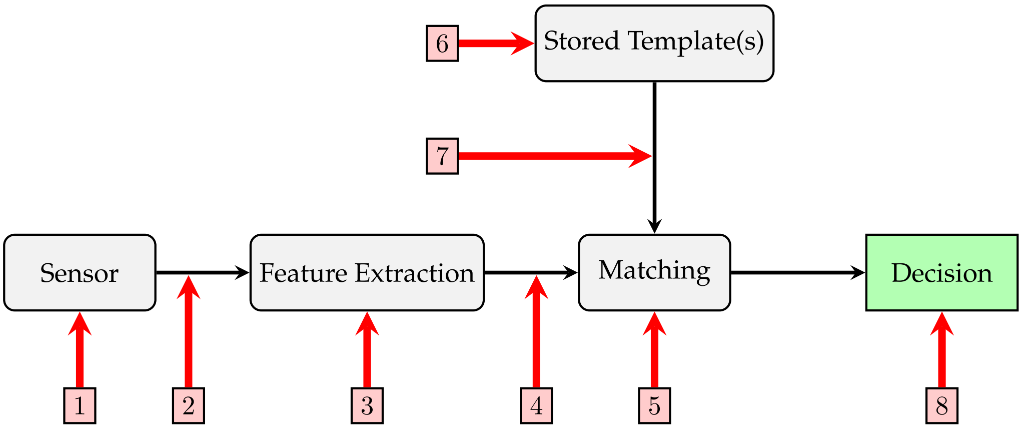


Figure 1.2.: Attack vectors in a generic biometric system.

standard[32] as “presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system”. The forged representations mimicking a genuine user are also denoted as spoofing artefacts. Some examples for such spoofing artefacts are fake fingerprint patterns using silicone [34], 3D face masks [29] and paper prints showing a face image [55], paper prints showing an image of an iris [65] as well as paper prints of vein patterns for finger- [69] and hand-vein recognition [68]. With a successful PA an impostor is able to gain illegitimate access to the system. To secure a biometric system against such threats, counter-measures to detect PAs can be put into place in the form of additional hard- or software performing presentation attack detection (PAD).

Digital image sensors are deployed in many biometric systems to acquire the desired biometric trait. Inserting a forged or even different biometric sample during the transmission between the biometric sensor and the feature extraction module (attack vector 2) is known as insertion attack. The inserted data could be acquired from a genuine user with another sensor off-site, even without his knowledge, or a manipulated image could be used to spoof the biometric system. Some biometric traits, such as fingerprints, irises, palm-prints, faces, etc. are inevitably presented to the open public and can easily be collected by the attacker, e.g. lifting fingerprints from a glass or taking a picture of the face of the desired subject with a telephoto lens where eventually also the iris can be extracted [38, 46].

Face recognition systems have experienced a major break-through with the recent developments in deep learning [63, 72], especially in unconstrained environments, thus leading to a significant improvement of the recognition performance due to the high generalisability of deep neural networks. However, this generalisability also made the systems more vulnerable to various attacks such as morphed face images [30]. Morphing techniques can be applied to resemble the biometric information of two or more individuals and create artificial biometric samples. In the case of a face morphing attack, face images from two subjects are combined using different morphing approaches, e.g. based on facial landmarks [30] or using deep learning techniques such as generative adversarial networks (GANs) [11]. These artificial samples are then successfully verified against probe samples of both individuals using state-of-the-art face recognition systems, if the morphed face image is used as reference image, i.e. as stored template (attack vector 6). A morphed (face) image might be introduced into the system during the enrolment process, for example when an individual applies for an electronic passport.

Realistic face morphs can be created by inexperienced users and non-experts with user-friendly applications [1]. Such realistic morphs are even able to fool human experts [31]. Several approaches have been proposed in literature to detect face morphing attacks, a recent overview can be found in [60].

The authenticity and integrity of the images processed within a biometric system plays an important role for its overall security, which could simply be achieved by encrypting the data. However, this may cause to replace the existing hardware because of the additional computational burden necessary for en- and decrypting the data. Biometric matching in the encrypted domain, might require to develop new algorithms, while other template protection approaches can lead to a recognition performance degradation as shown in [18], which might also not be feasible in every scenario. Watermarking of the data could also be employed, however this might as well introduce additional computational burden and could negatively impact the biometric system's performance, as shown for iris recognition in [39]. Alternatively, one could perform an analysis of the existing image data to detect anomalies. This is where Digital Image Forensics come into play, providing an extensive collection of tools for analysing images.

1.3. Digital Image Forensics

The field of Digital Image Forensics deals with structural analysis of image files and statistical analysis of the image data in order to investigate various traces in the digital images. Due to many biometric traits being acquired using digital image sensors and therefore being captured in digital images, it is intuitive to apply techniques from Digital Image Forensics for analysing specific properties of the images.

Digital Image Forensics aims at acquiring knowledge on visual contents and acquisition devices by evaluating the traces that are left on the data during the acquisition and in the subsequent processing. These intrinsic signatures are used to investigate different aspects, like identification of the source sensor of an image or video, or verification of the integrity of the data (if it has been modified or not) without any prior knowledge on it.

Every electronic device capable of acquiring digital images contains an imaging sensor. This sensor contains a large number of photo sensitive detectors made of silicon, commonly known as pixels. They have the ability to convert photons into electrons by exploiting the photoelectric effect [40, 43]. The charge accumulated in every pixel is first amplified and afterwards converted into a digital signal, which is further processed and then stored.

An important tool used to perform many forensic tasks is the photo-response non-uniformity (PRNU) of imaging sensors as described by Fridrich in [33]. It can be used for a variety of important tasks, such as device identification, device linking, recovery of processing history, and detection of digital forgeries. The PRNU is an intrinsic property of all digital imaging sensors, which is characterised by slight variations among individual pixels in their ability to convert photons to electrons, as illustrated in Figure 1.3. It shows an example of extracting the PRNU from an evenly illuminated part of an image and the corresponding PRNU, where the variation among the pixels can be observed.

Consequently, every sensor casts a weak noise-like pattern onto every image it takes. This pattern, which plays the role of a “sensor fingerprint”, is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering. This fingerprint can be estimated from images taken by the camera and later detected in a given image to establish the image's origin and integrity.

Even though the PRNU is stochastic in nature, it is a relatively stable component of the sensor over its life span and is therefore a very useful forensic quantity, responsible for a unique sensor

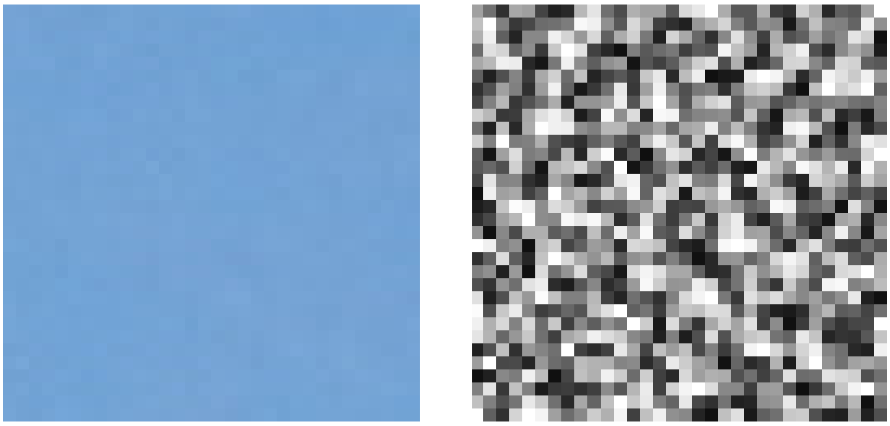


Figure 1.3.: Evenly illuminated image and extracted PRNU noise residual. The contrast of the noise residual has been enhanced to emphasise the variation among the pixels.

fingerprint with the following important properties [33]: dimensionality, universality, generality, stability and robustness. The PRNU fingerprint can be used for various forensic tasks [33]:

- By testing the presence of a specific fingerprint in the image, one can achieve reliable device identification (e.g., prove that a certain camera took a given image) or prove that two images were taken by the same device (device linking). The presence of camera fingerprint in an image is also indicative of the fact that the image under investigation is natural and not a computer rendering.
- By establishing the absence of the fingerprint in individual image regions, it is possible to discover replaced parts of the image (integrity verification).
- By detecting the strength or form of the fingerprint, it is possible to reconstruct some of the processing history. For example, one can use the fingerprint as a template to estimate geometrical processing, such as scaling, cropping, or rotation. Non-geometrical operations will also influence the strength of the fingerprint in the image and thus can potentially be detected.
- The spectral and spatial characteristics of the fingerprint can be used to identify the camera model or distinguish between a scan and a digital camera image (the scan will exhibit spatial anisotropy).

Several other image characteristics have been investigated for device identification, e.g. auto-white balance approximation [26], lens radial distortion [10], chromatic aberration, purple fringing, and sensor dust characteristics [27, 28]. Though, these approaches have not been investigated as extensively as the PRNU. An overview of image characteristics used for device identification is given in [42, 54, 57, 64, 8] and [52] gives an overview of forensic applications using the PRNU. Additionally, important properties as required for the deployment in a biometric system have been demonstrated. They comprise the management of large datasets [37, 36] and the robustness against common signal processing operations like compression and malicious signal processing [2, 58].

Device clustering can be seen as an extension of device linking, where images acquired with the same device are grouped into clusters. In this scenario, in contrast to the device identification one, a large set of images from unknown source cameras is investigated. The number of different devices as well as the distribution of the images among them is usually unknown. Furthermore, the acquisition of additional data is not possible because the source camera(s) are not available. As mentioned by Liu *et al.* [5] the varying number of images for each source camera could be challenging to supervised learning based approaches. Unbalanced category distribution has long been a problem in machine learning, where most classification algorithms generate classifiers by minimising the overall error rate, leading to either ignoring the minority class or over-fitting the training sample into many minor classes [44]. Several classical clustering techniques have been proposed in literature to solve this problem [47, 9, 3, 7, 50, 48, 49, 53] relying on the PRNU and utilising the same extraction procedures as for the device identification case. In order to find the "optimal" clustering solution, a cluster validity assessment [71] of the clustering outcome has to be conducted. In general, this assessment is based on two criteria:

- *Compactness*: the members of each cluster should be as close to each other as possible.
- *Separation*: the clusters themselves should be widely separated.

Therefore, the partition that best fits the underlying data can be considered as the "optimal" clustering solution. Various internal and external clustering validity indices (CVIs) have been proposed in literature [67]:

- *External CVIs*: measure of agreement between two clustering solutions where one is an a priori known reference.
- *Internal CVIs*: measure of quality by means of inherent quantities and features contained in the clustered data without use of external knowledge.

In the context of biometric systems the deduced sensor information can serve as a basis for various forensic and non-forensic applications, e.g. securing a biometric system against insertion attacks, enabling device selective processing of the image data or detection of image manipulations. Therefore, securing a system against insertion attacks requires to determine the origin of processed images and to ensure that only images acquired with the deployed biometric sensor are further processed. On the other hand, selective processing of the biometric images helps to improve the interoperability by applying a sensor tailored biometric tool chain [4]. For this purpose, information about the sensor model is required, which can be deduced from the images directly utilising image forensic methods [51]. Determining an image's origin can be done at different levels: sensor technology, brand, model, unit. While the former task (device identification) requires the identification to be performed at unit level, the latter one (selective processing) only requires to determine the device model because of the image processing being usually the same for all units of a specific device model.

2. Contribution

The work published in the scope of this thesis can be divided into three major areas: biometric sensor forensics, PRNU-based face morph detection and mobile applications.

2.1. Biometric Sensor Forensics

The identification of an image's origin enables different applications in a biometric system, such as enhancing its security by identifying the acquisition device or improvement of a system's performance in cases where sensors from different manufacturers and models are deployed. Previous work investigated the discrimination of biometric sensors on unit level in [70], [45] and [6] by making use of the PRNU. The outcome of Höller *et al.* [70] and Kalka *et al.* [45], where the discriminative power of iris sensors has been evaluated, showed strong variations in the identification performance for the investigated iris sensors. Bartlow *et al.* [6] investigated various fingerprint sensors and reported varying identification accuracy for the sensors as well. The discrimination of biometric sensors can be considered to be more challenging compared to common camera devices due to the nature of the acquired images. The image content of biometric data has a much higher correlation, thus making the quality of the extracted PRNU more vital to avoid false matching due to the correlated image content.

2.1.1. Source Sensor Clustering

Previous work [23] applied different clustering techniques to the CASIA-Iris V4 database in order to detect the presence of images from multiple sensors, where the results indicated that multiple sensors might have been used within each subset of the dataset. Following this lead, different PRNU enhancement and clustering techniques [22, 24, 25] have been evaluated on iris and fingerprint data. We observed that most PRNU enhancements did indeed improve the clustering results by increasing the differentiability of the various sensors' PRNU noise residuals, however the performance of the investigated combinations was highly situational as no single enhancement or clustering technique or combination of both was able to improve the clustering performance for all datasets alike.

In contrast to the previous contributions, our investigation in [19] was not conducted on biometric data. Instead it focused on the examination of a real word criminal case dataset containing still images found on a suspect's computer during a sexual abuse case investigation. The data was evaluated using different PRNU-based clustering algorithms and a quantitative analysis of the clustering results has been performed using various cluster validity indices (CVIs). Before analysing the criminal case data, a sanity check of all clustering techniques and validity indices has been conducted using data with known ground truth [35], which revealed the inability of certain algorithms and CVIs to correctly cluster the data and quantify the clustering output. The gained insights guided the evaluation of the criminal case data and supported giving some recommendations on how to apply PRNU-based clustering in such scenarios.

2.1.2. Sensor Identification

In the context of device identification, in [16] and [17] we examined visible spectrum and near-infrared (NIR) iris/ocular images to demonstrate whether an iris image's origin can be reliably determined among a large number of iris datasets. We discussed the applicability from a forensic and non-forensic perspective. Three different techniques have been applied in the evaluation: a photo response non-uniformity based one (PSI), an image texture feature based one (ITC) and finally the fusion of both PSI and ITC. The experiments included 19 iris datasets acquired with different sensors as well as a second dataset containing natural images from six different camera models with five instances each in order to investigate a discrimination on unit level as well as model level. The PSI approach was able to differentiate the sensors at unit level, but requires a certain minimum patch size. The ITC approach worked reliably in determining the iris image source regardless of the patch size, but only at model (dataset) level. The fusion of both ITC and PSI finally allowed for a unit level discrimination for a wide range of sensors.

Publications (sorted chronologically)

[22] L. Debiasi and A. Uhl. Blind biometric source sensor recognition using advanced PRNU fingerprints. In *Proceedings of the 2015 European Signal Processing Conference (EUSIPCO'15)*, Nice, France, 2015

[24] L. Debiasi and A. Uhl. Comparison of PRNU enhancement techniques to generate PRNU fingerprints for biometric source sensor attribution. In *Proceedings of the 4th International Workshop on Biometrics and Forensics (IWBF'16)*, Limassol, Cyprus, 2016

[25] L. Debiasi and A. Uhl. PRNU enhancement effects on biometric source sensor attribution. *IET Biometrics*, 4(6):256–265, 2017

[16] L. Debiasi, C. Kauba, and A. Uhl. Identifying iris sensors from iris images. In C. Rathgeb and C. Busch, editors, *Iris and Periocular Biometric Recognition*, chapter 16, pages 359–382. IET, London, UK, 2017

[17] L. Debiasi, C. Kauba, and A. Uhl. Identifying the origin of iris images based on fusion of local image descriptors and PRNU based techniques. In *Proceedings of the IAPR/IEEE International Joint Conference on Biometrics (IJCB'17)*, Denver, Colorado, USA, 2017

[19] L. Debiasi, E. Leitet, K. Norell, T. Tachos, and A. Uhl. Blind source camera clustering of criminal case data. In *Proceedings of the 7th International Workshop on Biometrics and Forensics (IWBF'19)*, Cancun, Mexico, 2019

2.2. Face Morph Detection

Morphed face images can be considered as a serious threat to face recognition systems by compromising the unique link between the biometric reference data and the associated subject. The ePassport application process in many countries allows for the applicant to provide a face image in digital or analogue form. Thus, in a face morphing attack scenario a criminal might end up with a valid ePassport retaining all document security features, but containing a morphed face image allowing successful verification of multiple individuals. A political activist successfully demonstrated how such a manipulated ePassport can be obtained in Germany[1]. Thus it is crucial to be able to detect morphing attacks in general and morphed face images in particular. An overview of proposed morph detection approaches is given in [60], where also general challenges and issues have been outlined consisting of morphed image quality, comparability and

[1] `https://www.vice.com/en_us/article/pa9vyb/peng-collective-artists-hack-german-passport`

reporting of results, robustness of morph detectors and the missing investigation of printed and re-scanned images.

In [21], we proposed a PRNU-based morph detection algorithm that analyses the changes within the PRNU caused by the morphing process in the spectral domain. The results on a comprehensive database of 961 bona fide and 2414 morphed face images showed a practical detection performance and also robustness of the algorithm to various image post-processing procedures such as sharpening or scaling, which might be applied to conceal the morphing process.

The PRNU-based morph detection algorithm was extended in [20] with a variance analysis of PRNU-based features across multiple image cells thus analysing relative changes among different image regions. This variability analysis helped in further improving the robustness to a wider range of post-processing procedures such as contrast enhancement via histogram equalisation.

Our work in [59] represents a significant extension of the two previous studies on PRNU-based morphing attack detection [21, 20]. The proposed system has been complemented by further investigations of different features in the spectral and spatial domain. Furthermore, the robustness of the system has been tested using four different morphing algorithms combined with a cross-database analysis and benchmarked against other state-of-the-art morph detection systems. The generalisability of the system across different cameras has also been verified on a dataset containing images acquired with 63 distinct camera instances (20 different models) across many camera manufacturers. A preliminary study on a dataset of printed and scanned images has also been included in the evaluation. Also, a vulnerability analysis of the proposed system was given with respect to attacks on the proposed morph detection system. In scenarios where the image source and morphing techniques are unknown, the proposed detector is shown to significantly outperform other previously established morphing attack detectors.

The focus of [13] lied on an experimental evaluation of the capabilities of various state-of-the-art morph detectors when confronted with classical landmark based morphing attacks (LMA) as well as a recently presented face morphing approach based on generative adversarial networks (MorGAN). The morph detection algorithms have been confronted with different attack scenarios consisting of known and unknown attacks with different morph types. In addition, the image quality of the morphed face images has been compared between LMA and MorGAN morphs. All investigated morph detection systems failed at consistently detecting all attacks, however the PRNU-based detection system proposed in [20] showed the most robust results although not being the best performing one.

Publications (sorted chronologically)

[21] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch. PRNU-based detection of morphed face images. In *Proceedings of the 6th International Workshop on Biometrics and Forensics (IWBF'18)*, Sassari, Italy, 2018

[20] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch. PRNU variance analysis for morphed face image detection. In *Proceedings of the IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS'18)*, Los Angeles, California, USA, 2018

[13] L. Debiasi, N. Damer, A. M. Saladie, C. Rathgeb, U. Scherhag, C. Busch, F. Kirchbuchner, and A. Uhl. On the detection of GAN-based face morphs using established morph detectors. In *Proceedings of the 20th International Conference on Image Analysis and Processing (ICIAP'19)*, Trento, Italy, 2019

[59] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl. Detection of face morphing attacks based on PRNU analysis. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 1(4):302–317, 2019

2.3. Mobile Applications

Mobile devices, smartphones in particular, are becoming more powerful with each generation and integrate a growing number of sensors: from RGB and in some cases NIR cameras, microphones, gyroscopes, accelerometers and GPS modules to dedicated biometric scanning hardware such as fingerprint readers and more. This hardware combined with growing computational capability in the form of more powerful CPUs and dedicated machine learning chips enables a wide range of applications using mobile devices like biometric authentication, health and government services, online banking, mobile ticketing, payment services or augmented reality applications just to name a few. Therefore these devices play an important role in modern everyday life and are used more and more frequently.

With the widespread use of these kind of devices, security concerns arise especially when they are used for government services or online banking. Notably, the biometric authentication use case has seen many (mostly successful) attack attempts like presentation attacks on mobile fingerprint[2], iris recognition[3] and face recognition systems[4].

To prevent such attacks, presentation attack detection (PAD) systems are employed on the mobile devices. In [41] we evaluated several mobile face recognition systems focussing on their usability and security. The usability was determined in form of successful authentication attempts by a genuine user und different environmental conditions, while the security was assessed by investigating the ability to detect presentation attacks of different complexity ranging from prints of a face on paper, replay attacks on screens to 3D masks. We observed that in general a higher security was only achieved at the cost of usability, going as far as making the systems more or less unusable for a genuine user.

As mentioned above, many biometric systems using external biometric traits (such as fingerprints, iris or face) have been already broken and deemed insecure in the context of mobile authentication. Therefore, alternatives for biometric authentication using traits that are not exposed externally are gaining attention. An example for such traits is vascular pattern recognition, which has been subject to interest in the scientific community and has already found its way into commercially available products in form of the LG G8 smartphone offering hand-vein recognition[5]. In [14], we designed and constructed a NIR illumination prototype add-on for smartphones to allow for mobile capturing of hand-veins for authentication purposes. In addition, to prevent fraudulent authentication attempts, a challenge response approach based on illumination variations was developed to ensure the authenticity of the acquired data. We acquired a hand-vein dataset containing images of 31 subjects from palmar and dorsal perspective of the hand and evaluated the recognition capabilities of different well-established vein recognition schemes on this challenging dataset.

An alternative application of mobile devices has been investigated in [15, 61, 62] by means of counterfeit drug detection. More specifically, we analysed the material structure of the drug packaging (paper and blister) in images acquired with a smartphone. The developed authentication system was able to detect intrinsic texture features of the packaging material without

[2] `https://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid`

[3] `https://www.ccc.de/en/updates/2017/iriden`

[4] `https://www.engadget.com/2019/12/16/facial-recognition-fooled-masks/`

[5] `https://www.businessinsider.de/lg-g8-smartphone-unlocks-with-hand-id-vein-palm-recognition-2019-2`

the use of additional physical markers. We assumed, that an impostor does not have access to the original drug manufacturer's packaging production facilities, thus he needs to produce the packaging using own materials such as cardboard for the outer packaging and plastics and metal foil for the blisters. The obtained results confirmed that the textural features of the drug packaging materials are constant and highly discriminative. The experiments furthermore indicated the possibility to train a classifier using a set of known instances that is able to authenticate unseen instances.

Publications (sorted chronologically)

[15] L. Debiasi, C. Kauba, R. Schraml, and A. Uhl. Towards drug counterfeit detection using package paperboard classification. In *Advances in Multimedia Information Processing – Proceedings of the 17th Pacific-Rim Conference on Multimedia (PCM'16)*, Springer LNCS, Xi'an, CHINA, 2016

[61] R. Schraml, L. Debiasi, C. Kauba, and A. Uhl. On the feasibility of classification-based product package authentication. In *IEEE Workshop on Information Forensics and Security (WIFS'17)*, Rennes, France, December 2017

[62] R. Schraml, L. Debiasi, and A. Uhl. Real or fake: Mobile device drug packaging authentication. In *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'18)*, Innsbruck, Austria, 2018

[14] L. Debiasi, C. Kauba, B. Prommegger, and A. Uhl. Near-infrared illumination add-on for mobile hand-vein acquisition. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS'18)*, Los Angeles, California, USA, 2018

[41] H. Hofbauer, L. Debiasi, and A. Uhl. Mobile face recognition systems: Exploring presentation attack vulnerability and usability. In *Proceedings of the 12th IAPR/IEEE International Conference on Biometrics (ICB'19)*, Crete, Greece, 2019

3. Publications

This chapter presents publications as originally published, reprinted with permission from the corresponding publishers. The publications are listed according to their order of mention within the publications lists in Chapter 2. The copyright of the original publications is held by the respective copyright holders, see the following copyright notices. In order to fit the paper dimension, reprinted publications may be scaled in size and/or cropped.

[22, 24, 17, 61, 21, 14, 20, 19, 41, 59] © 2015-2019 IEEE. Reprinted with permission. The original publications are available at IEEE Xplore Digital Library
`http://ieeexplore.ieee.org`

[15, 13] © 2016-2019 Springer. With kind permission from Springer Science+Business Media. The original publications are available at SpringerLink
`http://www.springerlink.com`

[25, 16] © 2017 IET. With kind permission from the Institution of Engineering and Technology. The original publications are available at the IET Digital Library
`http://digital-library.theiet.org`

[62] © 2018 ACM. With kind permission from the Association for Computing Machinery. The original publications are available at the ACM Digital Library
`http://dl.acm.org`

BLIND BIOMETRIC SOURCE SENSOR RECOGNITION USING ADVANCED PRNU FINGERPRINTS

Luca Debiasi, Andreas Uhl

Multimedia Signal Processing and Security Lab
University of Salzburg
Salzburg, Austria

ABSTRACT

Previous device identification studies on the iris sensors of the CASIA-Iris V4 database using PRNU fingerprints showed high variations regarding the differentiability of the sensors. These variations may have been caused by the usage of multiple sensors of the same model for the image acquisition. Since no specific documentation on this exists we investigate the presence of multiple image sensors in the data sets. The images under investigation, furthermore, show a strong correlation regarding their content, therefore we make use of different PRNU enhancements approaches based on weighting the PRNU depending on the image content. The enhanced PRNU is used in conjunction with different forensic techniques to detect the presence of multiple sensors in the data sets.

Finally, the results of the enhancement approaches and the results without any PRNU enhancement are compared and an assessment on whether multiple sensor instances have been used in the data sets is given.

Index Terms— Digital image forensics, Biometric sensor forensics, PRNU, Sensor identification

1. INTRODUCTION

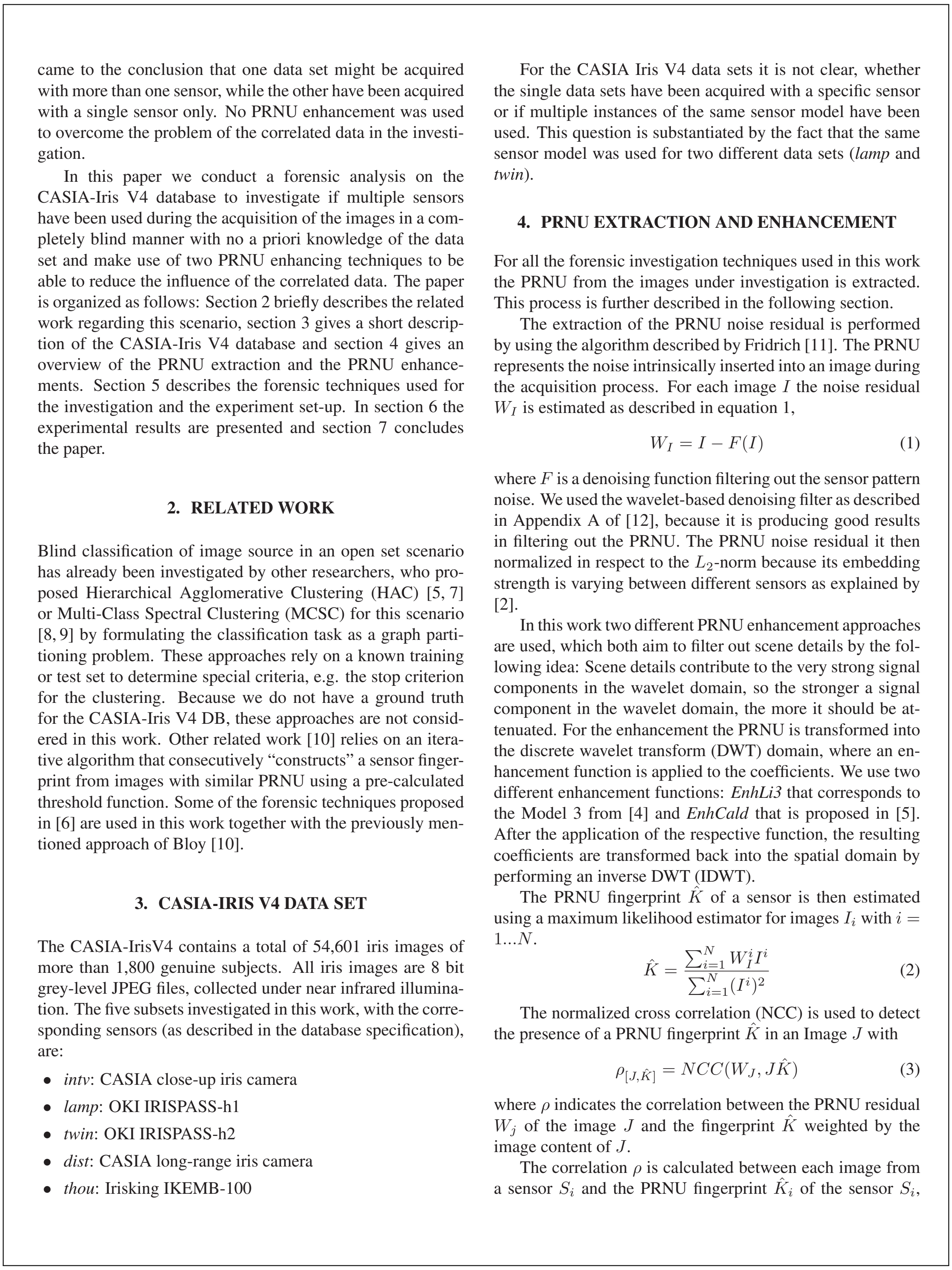
In the field of digital image forensics the photo response non-uniformity (PRNU) of an imaging sensor emerged as an important tool for the realization of different forensic tasks like device identification, device linking, recovery of processing history and the detection of digital forgeries.

Slight variations of individual pixels during the conversion of photons to electrons in digital image sensors are the source of the PRNU, thus it is considered an intrinsic property which is contained in all digital imaging sensors. Every digital image sensor adds this weak, noise-like pattern into every image that has been acquired with it. This pattern, which enables the identification of a specific image sensor, is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering and it meets essential criteria like dimensionality, universality, generality, stability and robustness [1] that make it well suited for forensic tasks.

Beyond that, the PRNU fingerprint of a sensor can also be used to improve a biometric systems security by ensuring the authenticity and integrity of images acquired with a biometric sensor. Previous studies on this application by Höller *et al.* [2] have conducted a feasibility study on the CASIA-Iris V4 database. The differentiability of the sensors in the CASIA-Iris V4 database using PRNU fingerprints has been tested with the conclusion, that the EERs and respective thresholds vary highly. Some sensors showed satisfying results while others did not, some subsets even showed EERs of over 20%. The question raised, that if PRNU fingerprints are being applied as an authentication measure for iris databases, it is not clear where the poor differentiation results for some sensors come from.

It was assumed that this high variation could be caused by the correlated data that was used to generate the sensors PRNU fingerprint. Further investigation from Debiasi *et al.* [3] showed that using uncorrelated data to generate the PRNU fingerprint does not improve the results for this data set and hence is not causing the high variation. An alternative method to deal with the uncorrelated data is to further separate the PRNU from the image content. Since the PRNU covers the high frequency components of an image, it is contaminated with other high frequency components from the images, such as edges. Li [4] proposed an approach for attenuating the influence of details from scenes on the PRNU so as to improve the device identification rate of the identifier. Caldelli *et al.* [5] considered this approach and developed a new kind of enhancer.

On the other hand, Höller *et al.* [2] suspected that multiple sensors may have been used for the acquisition of the CASIA Iris-V4 subsets. If a PRNU fingerprint is generated using images of different sensors, it will match with images acquired with all of these sensors and hence lead to a decreased differentiability. Unfortunately, neither the meta data of the images in the CASIA-Iris V4 database, nor the database description, denoting solely the sensor model without any additional information, can reveal the number of sensors instances used during the acquisition. Even the researchers involved in the acquisition cannot determine the number of sensors any more. Debiasi *et al.* [6] investigated the case of multiple sensors and



came to the conclusion that one data set might be acquired with more than one sensor, while the other have been acquired with a single sensor only. No PRNU enhancement was used to overcome the problem of the correlated data in the investigation.

In this paper we conduct a forensic analysis on the CASIA-Iris V4 database to investigate if multiple sensors have been used during the acquisition of the images in a completely blind manner with no a priori knowledge of the data set and make use of two PRNU enhancing techniques to be able to reduce the influence of the correlated data. The paper is organized as follows: Section 2 briefly describes the related work regarding this scenario, section 3 gives a short description of the CASIA-Iris V4 database and section 4 gives an overview of the PRNU extraction and the PRNU enhancements. Section 5 describes the forensic techniques used for the investigation and the experiment set-up. In section 6 the experimental results are presented and section 7 concludes the paper.

2. RELATED WORK

Blind classification of image source in an open set scenario has already been investigated by other researchers, who proposed Hierarchical Agglomerative Clustering (HAC) [5, 7] or Multi-Class Spectral Clustering (MCSC) for this scenario [8, 9] by formulating the classification task as a graph partitioning problem. These approaches rely on a known training or test set to determine special criteria, e.g. the stop criterion for the clustering. Because we do not have a ground truth for the CASIA-Iris V4 DB, these approaches are not considered in this work. Other related work [10] relies on an iterative algorithm that consecutively “constructs” a sensor fingerprint from images with similar PRNU using a pre-calculated threshold function. Some of the forensic techniques proposed in [6] are used in this work together with the previously mentioned approach of Bloy [10].

3. CASIA-IRIS V4 DATA SET

The CASIA-IrisV4 contains a total of 54,601 iris images of more than 1,800 genuine subjects. All iris images are 8 bit grey-level JPEG files, collected under near infrared illumination. The five subsets investigated in this work, with the corresponding sensors (as described in the database specification), are:

- *intv*: CASIA close-up iris camera
- *lamp*: OKI IRISPASS-h1
- *twin*: OKI IRISPASS-h2
- *dist*: CASIA long-range iris camera
- *thou*: Irisking IKEMB-100

For the CASIA Iris V4 data sets it is not clear, whether the single data sets have been acquired with a specific sensor or if multiple instances of the same sensor model have been used. This question is substantiated by the fact that the same sensor model was used for two different data sets (*lamp* and *twin*).

4. PRNU EXTRACTION AND ENHANCEMENT

For all the forensic investigation techniques used in this work the PRNU from the images under investigation is extracted. This process is further described in the following section.

The extraction of the PRNU noise residual is performed by using the algorithm described by Fridrich [11]. The PRNU represents the noise intrinsically inserted into an image during the acquisition process. For each image I the noise residual W_I is estimated as described in equation 1,

$$W_I = I - F(I) \tag{1}$$

where F is a denoising function filtering out the sensor pattern noise. We used the wavelet-based denoising filter as described in Appendix A of [12], because it is producing good results in filtering out the PRNU. The PRNU noise residual it then normalized in respect to the L_2-norm because its embedding strength is varying between different sensors as explained by [2].

In this work two different PRNU enhancement approaches are used, which both aim to filter out scene details by the following idea: Scene details contribute to the very strong signal components in the wavelet domain, so the stronger a signal component in the wavelet domain, the more it should be attenuated. For the enhancement the PRNU is transformed into the discrete wavelet transform (DWT) domain, where an enhancement function is applied to the coefficients. We use two different enhancement functions: *EnhLi3* that corresponds to the Model 3 from [4] and *EnhCald* that is proposed in [5]. After the application of the respective function, the resulting coefficients are transformed back into the spatial domain by performing an inverse DWT (IDWT).

The PRNU fingerprint $\hat{K}$ of a sensor is then estimated using a maximum likelihood estimator for images I_i with $i = 1...N$.

$$\hat{K} = \frac{\sum_{i=1}^{N} W_I^i I^i}{\sum_{i=1}^{N} (I^i)^2} \tag{2}$$

The normalized cross correlation (NCC) is used to detect the presence of a PRNU fingerprint $\hat{K}$ in an Image J with

$$\rho_{[J,\hat{K}]} = NCC(W_J, J\hat{K}) \tag{3}$$

where ρ indicates the correlation between the PRNU residual W_j of the image J and the fingerprint $\hat{K}$ weighted by the image content of J.

The correlation ρ is calculated between each image from a sensor S_i and the PRNU fingerprint $\hat{K}_i$ of the sensor S_i,

NoEnh	intv	lamp	twin	dist	thou
images	1307	6858	1095	1566	2000
partitions	143	212	20	1	6
partitions > 100	2	11	1	1	1
partitions < 10	128	157	18	0	4
unassociated images	0	0	0	0	0
EnhLi3	intv	lamp	twin	dist	thou
images	1307	6855	1095	1566	2000
partitions	186	266	24	1	14
partitions > 100	1	12	1	1	2
partitions < 10	168	129	19	0	12
unassociated images	0	0	0	0	0
EnhCald	intv	lamp	twin	dist	thou
images	1307	6855	1095	1566	2000
partitions	6	2867	307	1	193
partitions > 100	1	0	3	1	3
partitions < 10	4	260	254	0	188
unassociated images	928	0	0	0	0

Table 1: BFAIC experiment results on the CASIA-Iris V4 data sets for *NoEnh* (top), *EnhLi3* (middle) and *EnhCald* (bottom).

where only images are used that have not been part of the PRNU fingerprint estimation. Additionally the correlation ρ between all images from the other sensors S_j, $i \neq j$, and the PRNU fingerprint $\hat{K}_i$ of the sensor S_i is also calculated.

5. EXPERIMENTS AND SET-UP

All the subsets from the CASIA-Iris V4 DB are investigated independently. Since the image size is varying between the data sets, the PRNU noise residual of an image is extracted from 4 patches located in the corners with a size of 128x128 pixels each for all of the forensic techniques. Hence we obtain a total noise residual size of 256x256 pixels.

After the extraction of the PRNU noise residual, either no enhancement, the enhancement of Li [4] (denoted as *EnhLi3*) or the enhancement of Caldelli *et al.* [5] (denoted as *EnhCald*) is applied to the PRNU as described in section 4. A threshold value of $\alpha = 6$ was used for the enhancement function in both enhancement approaches.

After the extraction and optional enhancement, three different forensic techniques are applied to investigate the data sets:

- Blind Camera Fingerprinting and Image Clustering (*BCFAIC*) by Bloy [10]
- Sliding Window Fingerprinting (*SWF*) by Debiasi *et al.* [6]
- Device Identification on Dataset Partitions (*DIODP*) by Debiasi *et al.* [6]

6. RESULTS

In the following section the results of the ivestigation of the CASIA-Iris V4 DB with the before mentioned forensic techniques and PRNU enhancements are presented.

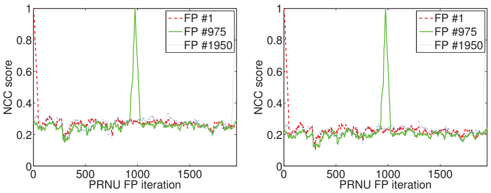


Fig. 1: Results of SW experiment *thou* data set without PRNU enhancement (left) and the *EnhLi3* enhancement (right).

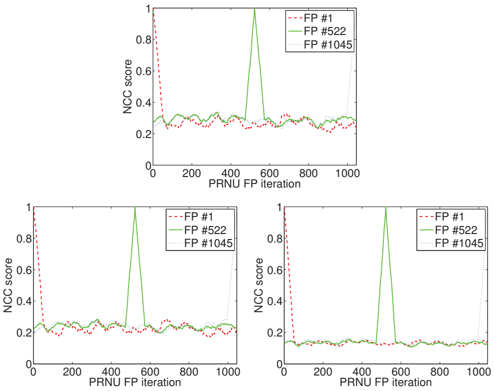


Fig. 2: Results of SW experiment for the *twin* data set with *NoEnh* (top), *EnhLi3* (left) and *EnhCald* (right).

6.1. Blind Camera Fingerprinting and Image Clustering

First the Blind Camera Fingerprinting and Image Clustering (BCFAIC) technique was applied to the different subsets of the CASIA-Iris V4 database. This technique creates clusters of associated images (images with a high NCC score) and partitions the data set. The resulting partitions should reflect the number of distinct sensors used in the data set. Unassociated images have a very low NCC score among each other, so that they are classified as being all from different sensors because they could not be clustered properly. Table 1 shows the results without any PRNU enhancement applied (*NoEnh*) as well as the results with the *EnhLi3* and *EnhCald* PRNU enhancement.

The results show a high cluster fragmentation for all subsets, except for the *dist* data set, where all images have been clustered together with all enhancement approaches. The *EnhLi3* enhancement produces slightly more clusters then *NoEnh*, but the results are comparable. The *EnhCald* enhancement, on the other hand, produces a much higher amount of clusters for all data sets (except *dist*) compared to the other enhancements and also leads to unassociated images in the *intv* data set.

The results of the BCFAIC experiments indicate that the *dist* dataset has been acquired with a single sensor, while the results are unclear for the other data sets. It can also be seen that the *EnhLi3* produces comparable results to the PRNU enhancement being omitted.

6.2. Sliding Window Fingerprinting

The Sliding Window Fingerprinting (SWP) moves a window with a defined size over the data image after image and a PRNU fingerprint from the data within this window is calculated in each step. The presence of images from multiple sensors in the data set should express in a sudden increase or decrease of the correlation score. If only images from one sensor are present in the data set, the correlation scores among all images should be quite stable around a certain level. The high spikes with a peak value of 1 occur when fingerprints that have one or more common images in their generation are compared.

As this experiment shows in figure 1, the *EnhLi3* enhancement produces comparable results as if no enhancement is applied for all data sets. There is only a very small offset in the correlation scores between the two configurations, where the *EnhLi3* scores are slightly lower, but the transitions are equal for both configurations. An example is given in figure 1 for the *thou* data set. Hence only the *EnhLi3* and *EnhCald* configurations are compared in the following figures.

In the results of the *dist*, *twin* and *thou* data sets no transitions in the correlation scores can be identified. They are comparable for *EnhLi3*, *EnhLi3* and *EnhCald*, therefore these data sets have probably been acquired with a single sensor according to this experiment. The only difference is an offset in the correlation scores for the individual enhancement configurations, as it can be seen in figure 2.

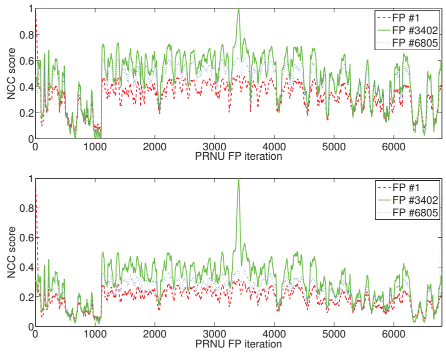


Fig. 3: Results of SW experiment for the *lamp* data set with *EnhLi3* (left) and *EnhCald* (right).

The figures 3 and 4 show the results for the *lamp* and *intv* data sets. In the *lamp* and *intv* data sets the previously described correlation score transitions can be observed at approximately iteration 700 and 1050 (*lamp*) and iteration 250 and 800 (*intv*).

Summing up, this technique suggests that all data sets, with the exception of *lamp* and *intv*, have been acquired with a single sensor. Regarding the PRNU enhancements it can be observed that the two PRNU enhancements *EnhLi3* and *EnhCald* exhibit decreased mean correlation scores.

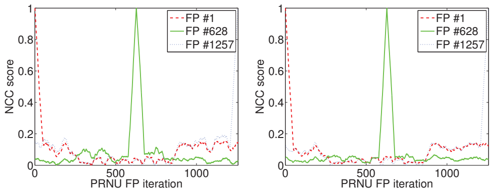


Fig. 4: Results of SW experiment for the *intv* data set with *EnhLi3* (left) and *EnhCald* (right).

6.3. Device Identification on Dataset Partitions

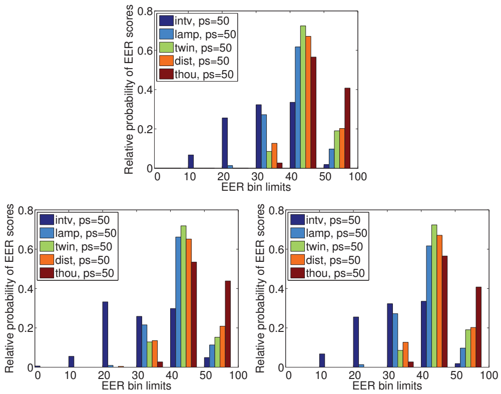


Fig. 5: Results of DIODP experiment on all CASIA-Iris V4 data sets with a partition size of 50.

The Device Identification on Dataset Partitions (DIODP) experiment divides the data sets into n partitions with the same size and treat the disjoint partitions as n different sensors. After calculating the pairwise EER scores for all partition combinations P_i and Pj, where $i \neq j$, the EER score distribution is evaluated. If the distribution contains mostly high EER scores, the data set probably contains images from a single sensor. On the other hand, if the distribution contains very low EER scores, the data set is suspicious of containing images from multiple sensors. To be able to clearly represent the resulting EER scores we performed a binning of the scores into six bins with the following limits: scores below 10%, between 10% and 20%, between 20% and 30%, between 30% and 40%, between 40% and 50%, and scores above 50%, where the lower bounds are inclusive and the upper bounds are exclusive.

Similar to the previous forensic techniques, the results for the two PRNU enhancement approaches are quite similar to the unenhanced ones, as represented in figure 5. This figure also indicates that the score distribution for the *intv* data set shows some low EER scores. For all other data sets it can be observed that the EER scores are mostly larger than 30%, which indicates that these data sets might be acquired with a single sensor. Having a closer look at the *intv* data set with different partition sizes in figure 6 indicates that this set might have been acquired with multiple sensors, because the distribution of the EER scores contains most of the scores in the range between 10% and 40% for almost all partition sizes under investigation.

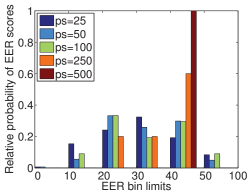


Fig. 6: Results of DIODP experiment with different partition sizes for the *intv* data set with the *EnhLi3*.

7. CONCLUSION

In this work we tried to establish a ground truth of the sensors used to acquire the various CASIA-Iris V4 data sets by using different PRNU enhancement techniques. This remains a challenging task for the CASIA-Iris V4 DB since this is a completely blind approach without any a priori knowledge of the sensors.

The PRNU enhancements did not clarify the previously obtained results from Debiasi *et al.* [6], where the results indicate that the *intv* data set might be acquired with more than sensor, while the other subsets have been acquired with one sensor. Actually, in this scenario, the impact of the evaluated PRNU enhancement approaches on the outcome of the applied forensic techniques is very low.

Unknown factors could have an impact on the quality of the PRNU noise residuals and hence tamper the results, therefore further studies have to be conducted to be able to use sensor fingerprints as an authentication measure for biometric systems.

8. ACKNOWLEDGMENTS

This work is partially funded by the Austrian Science Fund (FWF) under Project No. P26630.

REFERENCES

[1] Jessica Fridrich, “Sensor defects in digital image forensics,” in *Digital Image Forensics: There is more to a picture than meets the eye*, H.T. Sencar and N. Memon, Eds., chapter 6, pp. 179–218. Springer Verlag, 2012.

[2] Andreas Uhl and Yvonne Höller, “Iris-sensor authentication using camera PRNU fingerprints,” in *Proceedings of the 5th IAPR/IEEE International Conference on Biometrics (ICB’12)*, New Delhi, India, Mar. 2012, pp. 1–8.

[3] L. Debiasi, Z. Sun, and A. Uhl, “Generation of iris sensor PRNU fingerprints from uncorrelated data,” in *Proceedings of the 2nd International Workshop on Biometrics and Forensics (IWBF’14)*, 2014.

[4] Ch.-T. Li, “Source camera identification using enhanced sensor pattern noise,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, 2010.

[5] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti, “Fast image clustering of unknown source images,” in *IEEE International Workshop on Information Forensics and Security (WIFS) 2010*, 2010, pp. 1–5.

[6] L. Debiasi and A. Uhl, “Techniques for a forensic analysis of the casia-iris v4 database,” in *Proceedings of the 3rd International Workshop on Biometrics and Forensics (IWBF’14)*, 2015.

[7] Chang-Tsun Li, “Unsupervised classification of digital images using enhanced sensor pattern noise.,” in *ISCAS*. 2010, pp. 3429–3432, IEEE.

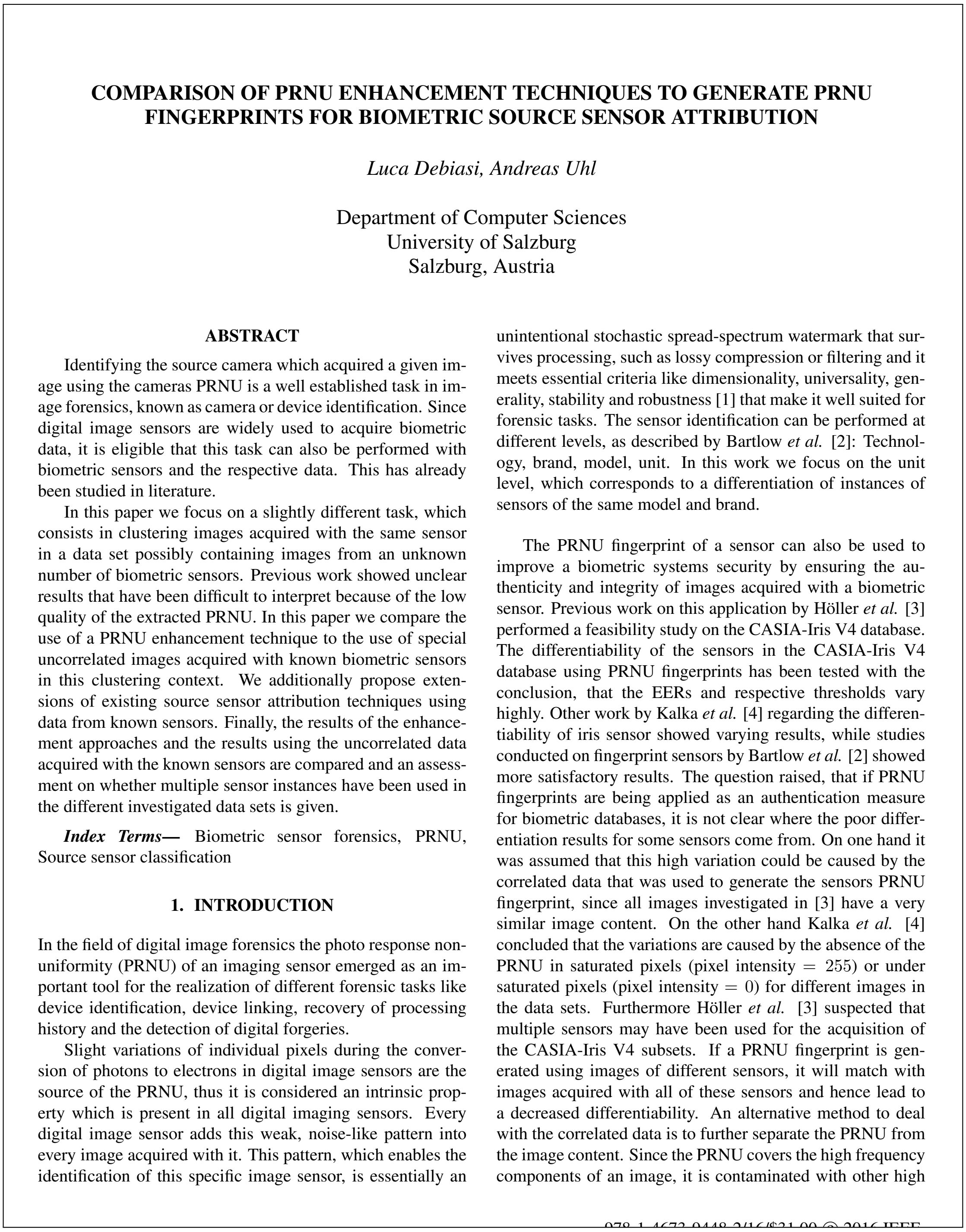
[8] Shuhan Luan, Xiangwei Kong, Bo Wang, Yanqing Guo, and Xingang You, “Silhouette coefficient based approach on cell-phone classification for unknown source images.,” in *ICC*. 2012, pp. 6744–6747, IEEE.

[9] I. Amerini, R. Caldelli, P. Crescenzi, A. Del Mastio, and A. Marino, “Blind image clustering based on the normalized cuts criterion for camera identification,” *Signal Processing: Image Communication*, , no. 29, pp. 831–843, 2014.

[10] G. Bloy, “Blind camera fingerprinting and image clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 532–534, Mar. 2008.

[11] J. Fridrich, “Digital image forensic using sensor noise,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, March 2009.

[12] Jan Lukas, Jessica J. Fridrich, and Miroslav Goljan, “Digital camera identification from sensor pattern noise.,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.



COMPARISON OF PRNU ENHANCEMENT TECHNIQUES TO GENERATE PRNU FINGERPRINTS FOR BIOMETRIC SOURCE SENSOR ATTRIBUTION

Luca Debiasi, Andreas Uhl

Department of Computer Sciences
University of Salzburg
Salzburg, Austria

ABSTRACT

Identifying the source camera which acquired a given image using the cameras PRNU is a well established task in image forensics, known as camera or device identification. Since digital image sensors are widely used to acquire biometric data, it is eligible that this task can also be performed with biometric sensors and the respective data. This has already been studied in literature.

In this paper we focus on a slightly different task, which consists in clustering images acquired with the same sensor in a data set possibly containing images from an unknown number of biometric sensors. Previous work showed unclear results that have been difficult to interpret because of the low quality of the extracted PRNU. In this paper we compare the use of a PRNU enhancement technique to the use of special uncorrelated images acquired with known biometric sensors in this clustering context. We additionally propose extensions of existing source sensor attribution techniques using data from known sensors. Finally, the results of the enhancement approaches and the results using the uncorrelated data acquired with the known sensors are compared and an assessment on whether multiple sensor instances have been used in the different investigated data sets is given.

***Index Terms*—** Biometric sensor forensics, PRNU, Source sensor classification

1. INTRODUCTION

In the field of digital image forensics the photo response non-uniformity (PRNU) of an imaging sensor emerged as an important tool for the realization of different forensic tasks like device identification, device linking, recovery of processing history and the detection of digital forgeries.

Slight variations of individual pixels during the conversion of photons to electrons in digital image sensors are the source of the PRNU, thus it is considered an intrinsic property which is present in all digital imaging sensors. Every digital image sensor adds this weak, noise-like pattern into every image acquired with it. This pattern, which enables the identification of this specific image sensor, is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering and it meets essential criteria like dimensionality, universality, generality, stability and robustness [1] that make it well suited for forensic tasks. The sensor identification can be performed at different levels, as described by Bartlow *et al.* [2]: Technology, brand, model, unit. In this work we focus on the unit level, which corresponds to a differentiation of instances of sensors of the same model and brand.

The PRNU fingerprint of a sensor can also be used to improve a biometric systems security by ensuring the authenticity and integrity of images acquired with a biometric sensor. Previous work on this application by Höller *et al.* [3] performed a feasibility study on the CASIA-Iris V4 database. The differentiability of the sensors in the CASIA-Iris V4 database using PRNU fingerprints has been tested with the conclusion, that the EERs and respective thresholds vary highly. Other work by Kalka *et al.* [4] regarding the differentiability of iris sensor showed varying results, while studies conducted on fingerprint sensors by Bartlow *et al.* [2] showed more satisfactory results. The question raised, that if PRNU fingerprints are being applied as an authentication measure for biometric databases, it is not clear where the poor differentiation results for some sensors come from. On one hand it was assumed that this high variation could be caused by the correlated data that was used to generate the sensors PRNU fingerprint, since all images investigated in [3] have a very similar image content. On the other hand Kalka *et al.* [4] concluded that the variations are caused by the absence of the PRNU in saturated pixels (pixel intensity $= 255$) or under saturated pixels (pixel intensity $= 0$) for different images in the data sets. Furthermore Höller *et al.* [3] suspected that multiple sensors may have been used for the acquisition of the CASIA-Iris V4 subsets. If a PRNU fingerprint is generated using images of different sensors, it will match with images acquired with all of these sensors and hence lead to a decreased differentiability. An alternative method to deal with the correlated data is to further separate the PRNU from the image content. Since the PRNU covers the high frequency components of an image, it is contaminated with other high frequency components from the images, such as edges. Li [5] proposed an approach for attenuating the influence of details from scenes on the PRNU so as to improve the device identification rate of the identifier.

In the previously described sensor identification task the PRNU fingerprints are usually pre-calculated using images from sensors available to the investigators. However this is not always the case in a realistic scenario, because the images under investigation could be part of an image set containing images from an unknown number of different cameras. Hence, before a source identification can be performed, images acquired with the same camera need to be identified and grouped together first. This task is known as source camera attribution in an open set scenario [6]. This has already been investigated by other researchers, who proposed Hierarchical Agglomerative Clustering (HAC) [7, 8] or Multi-Class Spectral Clustering (MCSC) for this scenario [6] by formulating the classification task as a graph partitioning problem. These approaches rely on a known training or test set to determine special criteria, e.g. the stop criterion for the clustering. Because the ground truth for the data sets is usually not available in this scenario, these approaches are not considered in this work. Other related work by Bloy [9] relies on an iterative algorithm that consecutively "constructs" a sensor fingerprint from images with similar PRNU using a pre-calculated threshold function. Some of the source sensor attribution techniques used in [10] are used in this work together with the previously mentioned approach of Bloy [9].

In this paper we perform a source sensor attribution on different biometric data sets to investigate if multiple sensors have been used during the acquisition of the images in a completely blind manner without a priori knowledge of the data sets described in Section 4. To enhance the quality of the extracted PRNU, we make use of a PRNU enhancing technique to be able to reduce the influence of the image content on the results as described in Section 2. Furthermore special uncorrelated data has been acquired with available sensors to generate PRNU fingerprints and the performance of using these fingerprints is compared to the use of the PRNU enhancement technique. To be able to use the uncorrelated data specifically acquired with the available sensors, alterations of the previously mentioned techniques used in [10] are proposed in Section 3. Section 5 explains the experimental set-up and in Section 6 the experimental results are presented. Finally Section 7 concludes the paper.

2. PRNU EXTRACTION AND ENHANCEMENT

The extraction of the PRNU noise residuals is performed by using the algorithm described by Fridrich [13]. For each image I the noise residual W_I is estimated as described in equation 1,

$$W_I = I - F(I) \tag{1}$$

where F is a denoising function filtering out the sensor pattern noise. We used the wavelet-based denoising filter as described in Appendix A of [14], because it is producing good results in filtering out the PRNU. The PRNU noise residual is then normalized in respect to the L_2-norm because its embedding strength is varying between different sensors as explained by [3]. As additional post processing steps a zero mean operation has been applied to each extracted PRNU noise residual to suppress artifacts with regular grid structure and a Wiener filtering is performed in the Discrete Fourier Transform (DFT) domain to suppress periodic artifacts in the calculated PRNU fingerprints.

In this work we apply a PRNU enhancement approach which aims at filtering out scene details using the following idea: Scene details contribute to the very strong signal components in the wavelet domain, so the stronger a signal component in the wavelet domain, the more it should be attenuated. For the enhancement the PRNU is transformed into the discrete wavelet transform (DWT) domain, where an enhancement function is applied to the coefficients. The enhancement function *ELi* used corresponds to the Model 3 proposed in [5]. After the application of the respective function, the resulting coefficients are transformed back into the spatial domain by performing an inverse DWT (IDWT).

The PRNU fingerprint $\hat{K}$ of a sensor is then estimated using a maximum likelihood estimator for images I_i with $i = 1...N$.

$$\hat{K} = \frac{\sum_{i=1}^{N} W_I^i I^i}{\sum_{i=1}^{N} (I^i)^2} \tag{2}$$

To enhance the PRNU fingerprints a Wiener filter is applied in the DFT domain, to suppress periodic artifacts as described in [1].

The peak correlation energy (PCE), as proposed in [1], is used to detect the presence of a PRNU fingerprint $\hat{K}$ in an Image I with

$$\rho_{[I,\hat{K}]} = PCE(W_i, I\hat{K}) \tag{3}$$

where ρ indicates the PCE score between the PRNU residual W_i of the image I and the fingerprint $\hat{K}$ weighted by the content of I.

3. SOURCE SENSOR ATTRIBUTION TECHNIQUES

For the source sensor attribution we use two different techniques: the *Blind Camera Fingerprinting and Image Clustering* (BCFAIC) proposed in [9] and the *Sliding Window Fingerprinting* (SWFP) proposed in [10]. Additionally we propose extensions of these methods for the case that the sensor is available to the investigators and uncorrelated data is used to generate the PRNU fingerprint (KSBCFAIC and KSSWFP), which are described in the following section. These is done by acquiring images with high saturation (but not over saturated) and smooth content, according to Fridrich [1]. The novel extensions of the existing methods are presented below.

3.1. KSBCFAIC

In [9] Bloy proposed the Blind Fingerprinting and Image Clustering (BFAIC) technique, which performs an agglomerative clustering to construct PRNU fingerprints from a mixed set of images, enabling identification of each images source camera without any prior knowledge of source. This technique solely depends on a pre-calculated threshold function. Using this threshold function t an automatic clustering algorithm performs the following steps:

1. Randomly select pairs of images until a pair is found whose noise correlation exceeds $t(1)$; average the PRNU of this pair to form a fingerprint.
2. Perform the first pass: for each remaining image, correlate the PRNU with the fingerprint. When the correlation value exceeds t(# of images in fingerprint cluster), average (cluster) it into the fingerprint. When $n = 50$ images have been averaged into the fingerprint or all images have been tried, stop and go to Step 3.
3. Perform the second pass: loop over all the unclustered images a second time, correlating with the current fingerprint and adding those that exceed the threshold. (Do not average more than 50 images into the fingerprint but allow more than 50 to be associated with the fingerprint.)
4. Repeat Step 1. Give up when Step 1 has tried 1000 pairs without success.

To be able to use the uncorrelated data, the first step (Step 1) is modified so that in the first iteration a PRNU Fingerprint is calculated from the uncorrelated data and the selection of two random images is skipped. After that each remaining image is correlated to this fingerprint as described in Step 2 and 3. After correlating all images, Step 1 is repeated as in the original algorithm by selecting two random images. We call this extension *Known Sensor Blind Camera Fingerprinting and Image Clustering* (KSBCFAIC).

3.2. KSSWFP

The Sliding Window Fingerprinting (SWFP) technique proposed in [11] consists of a so called "sliding window" with an arbitrary but fixed size n that moves over a data set image by image. This novel forensic technique uses an iterative algorithm which performs the following steps:

1. Start at image with index $i = 0$.
2. Gather images inside the sliding window with size n, hence the images with index $i \dots i + n$.
3. Extract the PRNU noise residual for each image.
4. Compute a PRNU fingerprint using the images inside the window.
5. Increment the index i by 1.
6. Repeat step 2 until all the images have been used to calculate a PRNU fingerprint.

Moving the window over the whole data set yields a list of PRNU fingerprints, which have been computed using sequential overlapping windows. For a data set containing m images, $m - n$ PRNU fingerprints are generated. After generating the fingerprints, the similarity of a PRNU fingerprint FP_i from the iteration i with all other fingerprints FP_j where $i \neq j$ is computed by calculating the PCE score of each fingerprint pair. This leads to a similarity matrix with size $(m - n) \times (m - n)$ containing all the pairwise PCE scores.

For the *Known Sensor Sliding Window Fingerprinting* (KSSWFP) a PRNU fingerprint is calculated with the uncorrelated data and then its PCE score to all sequentially overlapping PRNU fingerprints generated from the data set under investigation is calculated, which leads to a $(m - n)$ sized vector. High PCE scores in this vector indicate that the current PRNU fingerprint matches to the known sensor used to generate the uncorrelated data.

Data set name	Sensor	Modality
casiaLamp	OKI Irispass-h	Iris
stsmH100_2009	Irisguard H100 IRT	Iris
stsmH100_2013	Irisguard H100 IRT	Iris
stsmIPH_2009	OKI Irispass-h	Iris
stsmIPH_2013	OKI Irispass-h	Iris
casiaFP	Digital Persona UrU4000	Fingerprint
stsmURU_1	Digital Persona UrU4000 #1	Fingerprint
stsmURU_2	Digital Persona UrU4000 #2	Fingerprint

Table 1: Data set name, sensor model and according biometric modality.

4. BIOMETRIC DATA SETS

The data sets used in this paper consist of images for two different biometric modalities, iris and fingerprints, and are illustrated in table 1. The *casiaLamp* data set corresponds to the *CASIA-Iris-Lamp* data set present in the CASIA-Iris V4 database [1]. The *casiaFP* data set corresponds to the CASIA Fingerprint V5 database [1]. The remaining data sets have not been published, however the iris and fingerprint data sets starting with "stsm" and ending with "2013" have been acquired during a COST STSM as described in [12], while data sets ending with "2009" have been provided by the host institution during the STSM. The ground truth on the number of sensor instances used for the acquisition is only known for the *stsmH100_2013*, *stsmIPH_2013*, *stsmURU_1* and *stsmURU_2* data sets, which consists of 1 sensor instance. For all other data sets only the sensor model is known, but not how many instances of this model have been used.

All images are 8 bit grey-level JPEG files. The iris data has been collected under near infrared illumination, while the fingerprint sensors used red LEDs. The uncorrelated data used in this work to acquire the PRNU fingerprints for the known sensors has been acquired according to [12] for the sensors: *OKI Irispass-h*, *Irisguard H100 IRT*, *Digital Persona*

[1] CASIA Iris Image Database and CASIA Fingerprint V5 Database, http://biometrics.idealtest.org/

BCFAIC *ELi*	casiaLamp	stsmH100_2009	stsmH100_2013	stsmIPH_2009	stsmIPH_2013	casiaFP	stsmURU_1	stsmURU_2
Images	16213	908	1451	1620	970	19958	1000	1000
Total partitions	7	2	1	3	3	3	2	2
Partitions > 500	2	1	1	1	1	1	1	1
Partitions < 10	3	1	0	1	1	1	1	0
KSBCFAIC	casiaLamp	stsmH100_2009	stsmH100_2013	stsmIPH_2009	stsmIPH_2013	casiaFP	stsmURU_1	stsmURU_2
Total partitions	8	2	1	3	3	3 / 3	2	2
Partitions > 500	3	1	1	1	1	2 / 2	1	1
Partitions < 10	4	0	0	1	1	0 / 0	0	0
KSBCFAIC *ELi*	casiaLamp	stsmH100_2009	stsmH100_2013	stsmIPH_2009	stsmIPH_2013	casiaFP	stsmURU_1	stsmURU_2
Total partitions	7	3	1	3	3	3 / 3	2	2
Partitions > 500	2	1	1	1	1	2 / 2	1	1
Partitions < 10	2	2	0	2	2	0 / 0	0	0

Table 2: Clustering Results of the BCFAIC technique with applied *ELi* PRNU enhancement (top) compared to the KSBCFAIC technique using uncorrelated data (middle) and a combination of the *ELi* PRNU enhancement and the use of uncorrelated data for KSBCFAIC (bottom).

UrU4000 #1 and *Digital Persona UrU4000 #2*. To obtain high-quality PRNU fingerprints according to Fridrich [1], images with uncorrelated content and high saturation have been acquired. *Irisguard H100 IRT* sensor had no built-in quality assessment for the acquired images, hence the uncorrelated could be acquired as desired. For all other sensors the quality assessment partially prevented to acquire such images.

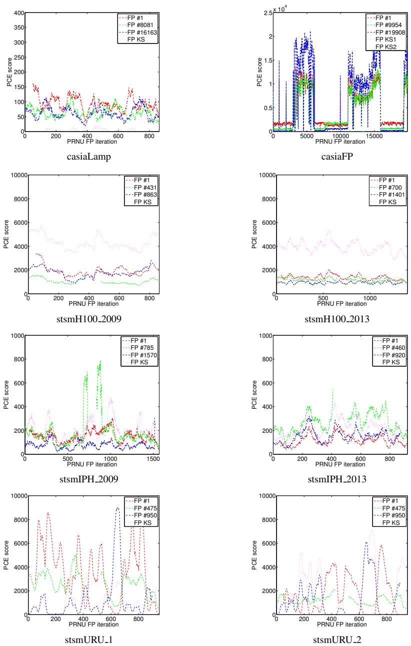
5. EXPERIMENTS AND SET-UP

All the data sets described in section 4 are investigated independently. Since the image size is varying between the data sets, the PRNU noise residual of each image is extracted from a single patch with a size of 256×256 pixels from the image centre. First we compare the use of PRNU enhancements for the ordinary source attribution techniques, BCFAIC and SWFP, to the extended techniques KSBCFAIC and KSSWFP without any further enhancements, to evaluate if the use of uncorrelated data helps to clarify the results for known sensors. Second, the use of PRNU enhancements and uncorrelated data are combined.

After the extraction of the PRNU noise residuals the enhancement of Li [5] (denoted as *ELi*) is applied to the PRNU as described in section 2. A threshold value of $\alpha = 6$ was used for the enhancement function for both enhancement approaches. The Wiener filtering in DFT is applied after each PRNU fingerprint calculation, while the zero mean operation is applied after the PRNU extraction for each image.

6. RESULTS

In the following section we first compare the use of PRNU enhancements for the ordinary source attribution techniques, BCFAIC and SWFP, to the extended techniques KSBCFAIC and KSSWFP without any further enhancements and after-



casiaLamp casiaFP

stsmH100_2009 stsmH100_2013

stsmIPH_2009 stsmIPH_2013

stsmURU_1 stsmURU_2

Fig. 1: Comparison of SWFP and KSSWFP for the various data sets: FP #x denotes the similarity of the PRNU fingerprint with iteration x to all other fingerprints for the SWFP technique, FP KS denotes the similarity of the PRNU fingerprint generated from uncorrelated data in the KSSWFP technique.

wards, we compare the results of using PRNU enhancements and uncorrelated data in combination to the previous results. For the *casiaFP* data set it was not clear which of the two sensors, *Digital Persona UrU4000 #1* or *Digital Persona UrU4000 #2*, has been used for the data acquisition, hence the uncorrelated data from both sensors was used independently for the experiments.

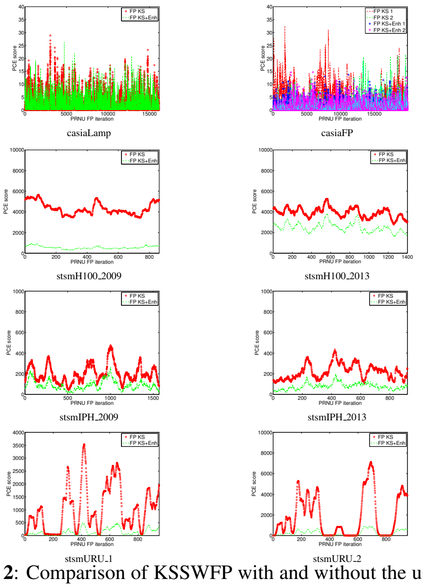
6.1. Uncorrelated data versus PRNU enhancements

First the Blind Camera Fingerprinting and Image Clustering (BCFAIC) using the *ELi* technique was applied to the different data sets and compared to the KSBCFAIC technique using data from the sensors assumed to have been used to acquire the data as shown in table 2. These techniques create clusters of associated images (images with a high PCE score) and partition the data sets. The resulting partitions are reflecting the number of distinct sensors used in the data set. The results do not show any clear improvement of using uncorrelated data for the sensors in respect to the PRNU enhancement, almost all data sets show one cluster containing almost all of the images and a small number of small clusters containing only a few images. Only the *casiaLamp* and *casiaFP* data sets show each two partitions both containing a large amount of images. This could be an indicator that the dataset is containing images from multiple sensors.

The (Known Sensor) Sliding Window Fingerprinting (KS)SWFP moves a window with a defined size over the data image after image and a PRNU fingerprint from the data within this window is calculated in each step. The presence of images from multiple sensors in the data set should express in a sudden increase or decrease of the correlation scores. If only images from one sensor are present in the data set, the correlation scores among all images should be quite stable around a certain level or have at least a PCE score of 50 or above. The results for the *casiaLamp* and *casiaFP* data set show many jumps in the PCE scores, which could indicate the presence of multiple sensors. It is to note that the PRNU fingerprint generated produces very low PCE scores (around 0), meaning that the uncorrelated data has not been acquired with the same sensor. The use of uncorrelated data does not lead to any improvement for the score interpretation here. Contrary to the previous results, for both *stsmH100* and both *stsmIPH* data sets, the results using the uncorrelated data show a high improvement in the PCE scores, which leads to the assumption that this sensor was exclusively used for their acquisition. For the remaining *stsmURU* data sets the scores have a high variation and the use of uncorrelated data does not help to clarify the scores either because the results are very similar to the PRNU enhancement results.

6.2. Combination of Uncorrelated data and PRNU enhancement

As it can be seen in table 2, the results of combining the *ELi* PRNU enhancement and uncorrelated data acquired also used to acquire the images in the respective data sets almost does not change the results for the KSBCFAIC technique. The only change that can be seen is that it shifts some images from the clusters containing between 500 and 10 images towards the larger clusters, which leads to a higher amount of small clusters with less than 10 images in some cases.



casiaLamp

casiaFP

stsmH100_2009

stsmH100_2013

stsmIPH_2009

stsmIPH_2013

stsmURU_1

stsmURU_2

Fig. 2: Comparison of KSSWFP with and without the use of the *ELi* enhancement: FP KS denotes the use of uncorrelated data only, while FP KS+Enh denotes the additional use of the *ELi* enhancement. The two numerations for *casiaFP* denote the sensors for the uncorrelated data: 1 - UrU4000 #1, 2 - UrU4000 #2.

The results of the combination for the KSSWFP technique show highly variable very low PCE scores for the *casiaLamp* and *casiaFP* data sets, from which no conclusion on the number of sensors can be made. The only assessment that can be done is that the uncorrelated data must have been acquired with a different sensor than the images in the data sets, since the PCE scores are all very low. For the *stsmH100_2009* the PCE scores drop quite drastically after the combination, but they remain at a level where one could state that the images in the data set have been acquired using the sensor to acquire the uncorrelated data. Both *stsmIPH* data sets and also the *stsmH100_2013* also show a decrease in the PCE scores, but not as radical as in the *stsmH100_2009* data set. The *stsmURU* data sets also show variable PCE scores ranging from very low to very high values, which implies the presence of multiple sensors. Since both data sets have been acquired using a single sensor instance only, as noted in section 4, the extracted PRNU must have a low quality and hence distort the results. The combination of PRNU enhancement and uncorrelated data also lowers the PCE scores for these two data sets.

7. CONCLUSION

In this paper we compared the use of PRNU enhancement techniques to the use of uncorrelated data for PRNU fingerprint generation in the context of sensor attribution. We investigated data from biometric sensors of two different biometric modalities, iris and fingerprint, where some of the data sets have been known to be acquired with a single sensor instance, while this was not known for others. We additionally proposed novel extension, KSBCFAIC and KSSWFP, for two existing source attribution techniques and compared them to the original techniques. Summing up the results of the comparison between PRNU enhancement and uncorrelated data, it can be stated that for some sensors, like the *OKI Irispass-h* and *Irisguard H100 IRT* iris sensors, the use of uncorrelated data improved the similarity between images of the data set and the PRNU fingerprint of the sensor, which shows up in the results of the KSSWFP technique. For the other data sets either the sensor was different than the one used to acquire the images in the data set (*casiaLamp*), or, especially for the fingerprint sensors, the extracted PRNU did not have a sufficient quality to ensure reliable results. Further studies have to be performed in this regard, since previous results from literature showed that the PRNU extracted for fingerprint sensors has a comparable quality to the one extracted from sensors of other biometric modalities.

8. ACKNOWLEDGMENTS

This work is partially funded by the Austrian Science Fund (FWF) under Project No. P26630.

9. REFERENCES

[1] Jessica Fridrich, “Sensor defects in digital image forensics,” in *Digital Image Forensics: There is more to a picture than meets the eye*, H.T. Sencar and N. Memon, Eds., chapter 6, pp. 179–218. Springer Verlag, 2012.

[2] N. Bartlow, N. Kalka, B. Cukic, and A. Ross, “Identifying sensors from fingerprint images,” in *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, June 2009, pp. 78–84.

[3] Andreas Uhl and Yvonne Höller, “Iris-sensor authentication using camera PRNU fingerprints,” in *Proceedings of the 5th IAPR/IEEE International Conference on Biometrics (ICB'12)*, New Delhi, India, Mar. 2012, pp. 1–8.

[4] Nathan Kalka, Nick Bartlow, Bojan Cukic, and Arun Ross, “A preliminary study on identifying sensors from iris images,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.

[5] Ch.-T. Li, “Source camera identification using enhanced sensor pattern noise,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, 2010.

[6] I. Amerini, R. Caldelli, P. Crescenzi, A. Del Mastio, and A. Marino, “Blind image clustering based on the normalized cuts criterion for camera identification,” *Signal Processing: Image Communication*, , no. 29, pp. 831–843, 2014.

[7] Chang-Tsun Li, “Unsupervised classification of digital images using enhanced sensor pattern noise.,” in *ISCAS*. 2010, pp. 3429–3432, IEEE.

[8] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti, “Fast image clustering of unknown source images,” in *IEEE International Workshop on Information Forensics and Security (WIFS) 2010*, 2010, pp. 1–5.

[9] G. Bloy, “Blind camera fingerprinting and image clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 532–534, Mar. 2008.

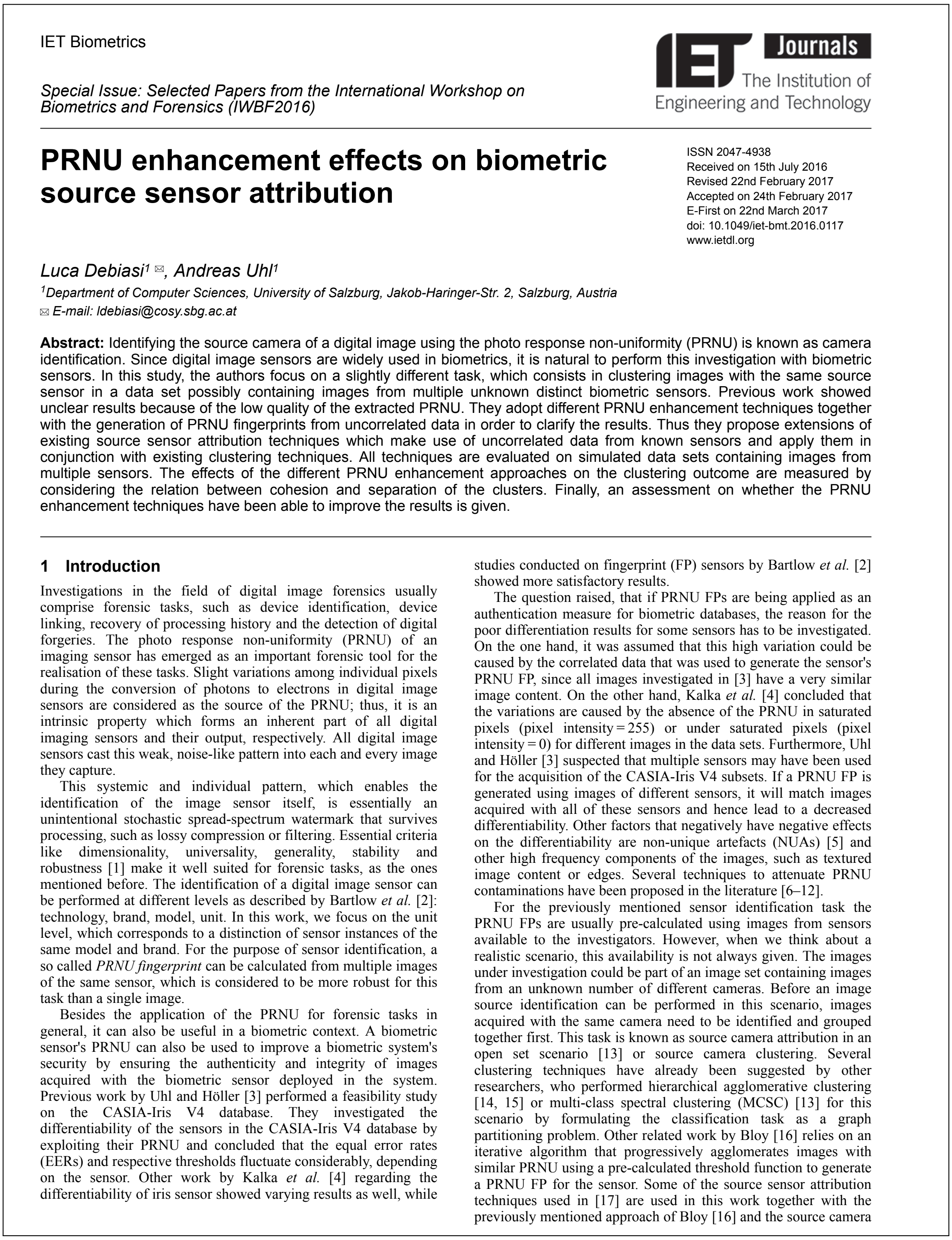
[10] L. Debiasi and A. Uhl, “Techniques for a forensic analysis of the casia-iris v4 database,” in *Proceedings of the 3rd International Workshop on Biometrics and Forensics (IWBF'15)*, 2015.

[11] L. Debiasi and A. Uhl, “Blind biometric source sensor recognition using advanced prnu fingerprints,” in *Proceedings of the 2015 European Signal Processing Conference (EUSIPCO 2015)*, 2015.

[12] L. Debiasi, Z. Sun, and A. Uhl, “Generation of iris sensor PRNU fingerprints from uncorrelated data,” in *Proceedings of the 2nd International Workshop on Biometrics and Forensics (IWBF'14)*, 2014.

[13] J. Fridrich, “Digital image forensic using sensor noise,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, March 2009.

[14] Jan Lukas, Jessica J. Fridrich, and Miroslav Goljan, “Digital camera identification from sensor pattern noise.,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.



IET Biometrics

Special Issue: Selected Papers from the International Workshop on Biometrics and Forensics (IWBF2016)

Journals
The Institution of Engineering and Technology

PRNU enhancement effects on biometric source sensor attribution

ISSN 2047-4938
Received on 15th July 2016
Revised 22nd February 2017
Accepted on 24th February 2017
E-First on 22nd March 2017
doi: 10.1049/iet-bmt.2016.0117
www.ietdl.org

Luca Debiasi[1] ✉, *Andreas Uhl*[1]

[1]*Department of Computer Sciences, University of Salzburg, Jakob-Haringer-Str. 2, Salzburg, Austria*

✉ *E-mail: ldebiasi@cosy.sbg.ac.at*

Abstract: Identifying the source camera of a digital image using the photo response non-uniformity (PRNU) is known as camera identification. Since digital image sensors are widely used in biometrics, it is natural to perform this investigation with biometric sensors. In this study, the authors focus on a slightly different task, which consists in clustering images with the same source sensor in a data set possibly containing images from multiple unknown distinct biometric sensors. Previous work showed unclear results because of the low quality of the extracted PRNU. They adopt different PRNU enhancement techniques together with the generation of PRNU fingerprints from uncorrelated data in order to clarify the results. Thus they propose extensions of existing source sensor attribution techniques which make use of uncorrelated data from known sensors and apply them in conjunction with existing clustering techniques. All techniques are evaluated on simulated data sets containing images from multiple sensors. The effects of the different PRNU enhancement approaches on the clustering outcome are measured by considering the relation between cohesion and separation of the clusters. Finally, an assessment on whether the PRNU enhancement techniques have been able to improve the results is given.

1 Introduction

Investigations in the field of digital image forensics usually comprise forensic tasks, such as device identification, device linking, recovery of processing history and the detection of digital forgeries. The photo response non-uniformity (PRNU) of an imaging sensor has emerged as an important forensic tool for the realisation of these tasks. Slight variations among individual pixels during the conversion of photons to electrons in digital image sensors are considered as the source of the PRNU; thus, it is an intrinsic property which forms an inherent part of all digital imaging sensors and their output, respectively. All digital image sensors cast this weak, noise-like pattern into each and every image they capture.

This systemic and individual pattern, which enables the identification of the image sensor itself, is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering. Essential criteria like dimensionality, universality, generality, stability and robustness [1] make it well suited for forensic tasks, as the ones mentioned before. The identification of a digital image sensor can be performed at different levels as described by Bartlow *et al.* [2]: technology, brand, model, unit. In this work, we focus on the unit level, which corresponds to a distinction of sensor instances of the same model and brand. For the purpose of sensor identification, a so called *PRNU fingerprint* can be calculated from multiple images of the same sensor, which is considered to be more robust for this task than a single image.

Besides the application of the PRNU for forensic tasks in general, it can also be useful in a biometric context. A biometric sensor's PRNU can also be used to improve a biometric system's security by ensuring the authenticity and integrity of images acquired with the biometric sensor deployed in the system. Previous work by Uhl and Höller [3] performed a feasibility study on the CASIA-Iris V4 database. They investigated the differentiability of the sensors in the CASIA-Iris V4 database by exploiting their PRNU and concluded that the equal error rates (EERs) and respective thresholds fluctuate considerably, depending on the sensor. Other work by Kalka *et al.* [4] regarding the differentiability of iris sensor showed varying results as well, while studies conducted on fingerprint (FP) sensors by Bartlow *et al.* [2] showed more satisfactory results.

The question raised, that if PRNU FPs are being applied as an authentication measure for biometric databases, the reason for the poor differentiation results for some sensors has to be investigated. On the one hand, it was assumed that this high variation could be caused by the correlated data that was used to generate the sensor's PRNU FP, since all images investigated in [3] have a very similar image content. On the other hand, Kalka *et al.* [4] concluded that the variations are caused by the absence of the PRNU in saturated pixels (pixel intensity = 255) or under saturated pixels (pixel intensity = 0) for different images in the data sets. Furthermore, Uhl and Höller [3] suspected that multiple sensors may have been used for the acquisition of the CASIA-Iris V4 subsets. If a PRNU FP is generated using images of different sensors, it will match images acquired with all of these sensors and hence lead to a decreased differentiability. Other factors that negatively have negative effects on the differentiability are non-unique artefacts (NUAs) [5] and other high frequency components of the images, such as textured image content or edges. Several techniques to attenuate PRNU contaminations have been proposed in the literature [6–12].

For the previously mentioned sensor identification task the PRNU FPs are usually pre-calculated using images from sensors available to the investigators. However, when we think about a realistic scenario, this availability is not always given. The images under investigation could be part of an image set containing images from an unknown number of different cameras. Before an image source identification can be performed in this scenario, images acquired with the same camera need to be identified and grouped together first. This task is known as source camera attribution in an open set scenario [13] or source camera clustering. Several clustering techniques have already been suggested by other researchers, who performed hierarchical agglomerative clustering [14, 15] or multi-class spectral clustering (MCSC) [13] for this scenario by formulating the classification task as a graph partitioning problem. Other related work by Bloy [16] relies on an iterative algorithm that progressively agglomerates images with similar PRNU using a pre-calculated threshold function to generate a PRNU FP for the sensor. Some of the source sensor attribution techniques used in [17] are used in this work together with the previously mentioned approach of Bloy [16] and the source camera

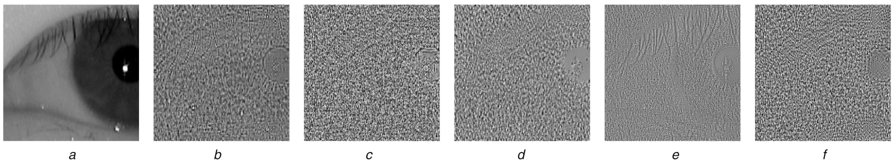


Fig. 1 *Comparison of different denoising filters and PRNU enhancements applied to a cropped iris image from the H100_2013 data set* ***(a)*** *Original image,* ***(b)*** $F_{\text{Luk}} + \text{ZM}$, ***(c)***$F_{\text{Luk}} + \text{Li}$, ***(d)*** F_{BM3D}, ***(e)*** F_{FSTV}, ***(f)***$F_{\text{Mih}} + \text{FDR}$

attribution techniques proposed in [14, 15, 18]. The size of the extracted PRNU for consumer cameras used for source sensor attribution found in the literature ranges from a very small size of 128 × 128 [15], 256 × 512 [14], 640 × 480, [19] to full size images of several megapixels, where the most common size appears to be 1024 × 1024 [16, 20]. The results reported for consumer cameras show that the size of the extracted PRNU plays a major role for the performance of the various techniques, where plausible results can be obtained with PRNU patches larger than 1024 × 1024 pixels in general and 256 × 512 pixels using additional PRNU enhancements.

In this work, we conduct a source sensor attribution on different biometric data sets from different biometrics modalities, which aims at determining whether the images in the data sets described in Section 4 have been acquired using multiple instances of the same sensor model. The investigation is conducted without taking any a priori knowledge about the sensors into consideration. To improve the quality of the extracted PRNU, we make use of various PRNU enhancement techniques which aim at attenuating undesired artefacts in the extracted PRNU as described in Section 2. Furthermore, additional uncorrelated data acquired with the same sensors as utilised to acquire the data sets is used for the generation of high-quality PRNU FPs. The performance of using the high-quality PRNU FPs is compared to the application of the various PRNU enhancement techniques. We propose novel extensions of the previously mentioned source sensor attribution techniques in Section 3 to be able to make use of the uncorrelated data. Section 5 explains the experimental set-up and describes the measure used for the evaluation of the clustering outcome and also contains the discussion of the experimental results. Finally, Section 6 concludes the paper.

This work is an extended version of a paper previously published in [21]. We extend our previous work by proposing additional source sensor attribution techniques that make use of uncorrelated data from known sensors and measure their performance on simulated data sets containing images from multiple sensors and different PRNU sizes as well as on existing biometric data sets mostly containing an unknown number of source sensors. Furthermore a quantitative assessment on the effects of using data from known sensors compared to various PRNU enhancement approaches and the combination of both of them is given based on a metric measuring the cohesion and separation of the clustering result for each technique.

2 PRNU extraction and enhancement

The extraction of the PRNU noise residuals is performed by applying Fridrich's approach [22]. For each image $\boldsymbol{I}$ the noise residual $\boldsymbol{W_I}$ is estimated as described in the following equation:

$$\boldsymbol{W_I} = \boldsymbol{I} - F(\boldsymbol{I}) \tag{1}$$

where F is a denoising function filtering out the sensor pattern noise. In this work, we made use of four different denoising algorithms: The two wavelet-based denoising filters proposed by Lukas *et al.* in Appendix A of [23] (F_{Luk}) and Mihcak *et al.* in [24] (F_{Mih}), the BM3D denoising filter proposed by Dabov *et al.* [6] (F_{BM3D}) and the FSTV algorithm proposed by Gisolf *et al.* [9] (F_{FSTV}).

After the PRNU extraction the noise residual $\boldsymbol{W_I}$ may be contaminated with undesired artefacts. To attenuate their effects different PRNU enhancement techniques have been proposed in the literature. Zero-meaning of the noise residuals's pixel rows and columns (*ZM*) removes NUAs with regular grid structures as described in [22]. Li [7] developed a technique for attenuating the influence of scene details or textured image content on the PRNU so as to improve the device identification rate of the identifier. This approach is referred to as *Li*. According to Lin and Li [12] some components of the extracted PRNU noise residual are severely contaminated by the errors introduced by denoising filters. They proposed a filtering distortion removal (FDR) algorithm that improves the quality of $\boldsymbol{W_I}$ by abandoning those components. The extracted and enhanced PRNU noise residual for a sample image using the various denoising filters and PRNU enhancements can be seen in Fig. 1.

Finally, the PRNU noise residual $\boldsymbol{W_I}$ is normalised with respect to the L_2-norm because its embedding strength is varying between different sensors as explained by Uhl and Höller [3].

The PRNU FP $\hat{\boldsymbol{K}}$ of a sensor is then estimated using a maximum-likelihood estimator for images I_i with $i = 1, \ldots, N$.

$$\hat{\boldsymbol{K}} = \frac{\sum_{i=1}^{N} \boldsymbol{W}_{\boldsymbol{I}}^{i} I_i}{\sum_{i=1}^{N} (I_i)^2} \tag{2}$$

PRNU FPs can be contaminated with NUAs as well. To further enhance the quality of PRNU FPs a Wiener filtering (WF) applied in the discrete Fourier transform domain is proposed in [1] to suppress periodic artefacts. Lin and Li [11] proposed a novel scheme named spectrum equalisation algorithm (SEA), where the magnitude spectrum of the PRNU FP $\hat{\boldsymbol{K}}$ is equalised through detecting and suppressing the peaks according to the local characteristics, aiming at removing the interfering periodic artefacts.

A method to detect the presence of a specific PRNU FP in an image which has not been geometrically transformed is the normalised cross correlation (NCC), which is defined as

$$\text{NCC}(\boldsymbol{A}, \boldsymbol{B}) = \frac{\sum_{w=1}^{W} \sum_{h=1}^{H} (\boldsymbol{A}(w,h) - \bar{A})(\boldsymbol{B}(w,h) - \bar{B}))}{\sqrt{\left(\sum_{w=1}^{W} \sum_{h=1}^{H} (\boldsymbol{A}(w,h) - \bar{A})^2\right)\left(\sum_{w=1}^{W} \sum_{h=1}^{H} (\boldsymbol{B}(w,h) - \bar{B})^2\right)}} \tag{3}$$

$\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices of the same size $w \times h$ and $\bar{A}$ and $\bar{B}$ are their respective mean. The mean of a matrix $\boldsymbol{X}$ with size $w \times h$ is defined as

$$\bar{X} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \boldsymbol{X}(w,h) \tag{4}$$

The NCC is used to detect the presence of a PRNU FP $\hat{\boldsymbol{K}}$ in an image $\boldsymbol{I}$ with

$$\rho_{[\boldsymbol{I}, \hat{\boldsymbol{K}}]} = \text{NCC}(\boldsymbol{W_I}, \boldsymbol{I}\hat{\boldsymbol{K}}) \tag{5}$$

where ρ indicates the correlation between the noise residual $\boldsymbol{W_I}$ of the image $\boldsymbol{I}$ and the PRNU FP $\hat{\boldsymbol{K}}$ weighted by the image content of $\boldsymbol{I}$.

On the other hand, the NCC can also be used to measure the similarity of two PRNU noise residuals $\hat{\boldsymbol{W}}_{\boldsymbol{I}}$ and $\hat{\boldsymbol{W}}_{\boldsymbol{J}}$ from two sensors S_i and S_j, as shown in the following equation:

$$\rho_{[\hat{\boldsymbol{W}}_{\boldsymbol{I}}, \hat{\boldsymbol{W}}_{\boldsymbol{J}}]} = \mathrm{NCC}(\hat{\boldsymbol{W}}_{\boldsymbol{I}}, \hat{\boldsymbol{W}}_{\boldsymbol{J}}) \tag{6}$$

Fridrich [1] proposed an alternative technique for measuring the similarity of two PRNU noise residuals or a PRNU noise residual and a PRNU FP, the peak correlation energy (PCE), which has proven to be yield more stable results in a scenario where the images have been subject to geometrical transformations, such as rotations or scaling. Since all images used in this work have not undergone any of these transformations and Kang *et al.* showed that PCE by definition may increase the false positive rate in [25], we decided to use the NCC over the PCE.

3 Source sensor attribution techniques

In this work, we consider various techniques for the source sensor attribution task, where we apply various existing source attribution techniques and propose a novel one. We furthermore propose novel extensions for these existing methods for the case that the sensor is available to the investigators and uncorrelated data is used to generate the PRNU FP. The uncorrelated data is generated by acquiring images with high saturation (but not over saturated) and smooth content, according to Fridrich [1]. All the mentioned clustering techniques generate a list of clusters, where the association of each image in the investigated data set to a cluster and thus a cluster label is obtained. The novel extensions of the existing methods together with a brief explanation of the original techniques are given in the following section.

3.1 Known sensor blind camera fingerprinting and image clustering ((KS)BCF)

In [16] Bloy proposed the blind camera fingerprinting and image clustering (BCF) technique, which performs an agglomerative clustering to construct PRNU FPs from a mixed set of images, enabling identification of each image's source camera without any prior knowledge of source. This technique solely depends on a pre-calculated threshold function. Using this threshold function t an automatic clustering algorithm performs the following steps:

1. Randomly select pairs of images until a pair is found whose noise correlation exceeds $t(1)$; average the PRNU of this pair to form a FP.
2. Perform the first pass: for each remaining image, correlate the PRNU with the FP. When the correlation value exceeds t(# of images in FP cluster), average (cluster) it into the FP. When $n = 50$ images have been averaged into the FP or all images have been tried, stop and go to Step 3.
3. Perform the second pass: loop over all the unclustered images a second time, correlating with the current FP and adding those that exceed the threshold. (Do not average more than 50 images into the FP but allow more than 50 to be associated with the FP.)
4. Repeat Step 1. Stop when Step 1 has tried 1000 pairs without success.

To be able to use the uncorrelated data, the first step (Step 1) is modified so that during the first iteration a PRNU FP is calculated from the uncorrelated data and the selection of two random images is skipped. After this modified step each remaining image is compared to this FP as described in Steps 2 and 3. After comparing all images, Step 1 is repeated as in the original algorithm by selecting two random images. We call this extension *Known Sensor Blind Camera Fingerprinting and Image Clustering* (KSBCF), as noted in the original paper [21].

3.2 Known sensor sliding window fingerprinting ((KS)SWx)

The Sliding Window Fingerprinting (SW) technique proposed in [26] consists of a so called ‘sliding window’ with an arbitrary but fixed size n that moves over a data set image by image. This forensic technique uses an iterative algorithm which performs the following steps:

1. Start at image with index $i = 0$.
2. Gather images inside the sliding window with size n, hence the images with index $i, \ldots, i + n$.
3. Extract the PRNU noise residual for each image.
4. Compute a PRNU FP using the images inside the window.
5. Increment the index i by 1.
6. Repeat step 2 until all the images have been used to calculate a PRNU FP.

Moving the window over the whole data set yields a list of PRNU FPs, which have been computed using sequential overlapping windows. For a data set containing m images, $m - n$ PRNU FPs are generated. After generating the FPs, the similarity of a PRNU FP FP_i from the iteration i with all other FPs FP_j where $i \neq j$ is computed by calculating the NCC score of each FP pair. This leads to a similarity matrix with size $(m - n) \times (m - n)$ containing all the pairwise NCC scores. The NCC scores of the PRNU FP comparisons where the FPs contain at least one common image are set to 0 because their correlation score would be much higher than average and introduce a bias to the clustering.

In [26], the number of clusters is determined in an explorative way by observing changes of the correlation scores. This leads to a rather vague estimation of the cluster structure in the data set. Hence, to assess the underlying cluster structure in a quantitative manner, we propose to apply different existing clustering techniques to cluster the obtained similarity matrix of pairwise PRNU FP comparisons. In this work, we applied the unsupervised clustering of digital images (UCDIs) [14], the fast image clustering (FIC) [15] and finally the MCSC algorithm [18]. The lower case ‘x’ in the technique name indicates the applied clustering technique: *U* for UCDI, *F* for FIC and *M* for the MCSC technique.

These techniques yield a list of clusters and the PRNU FPs associated to each cluster. To obtain a cluster association for each image in the data set instead of each generated PRNU FP, we perform a majority voting based on the images used to generate each PRNU FP and the cluster association: Each image is used for the generation of multiple PRNU FPs because of the sliding window property, hence we count the cluster association frequency of the PRNU FPs, which contain the specific image, and select the highest cluster label occurrence as the final decision for the image. This gives a cluster label for each image in the data set.

For the *Known Sensor Sliding Window Fingerprinting* (KSSWx) a PRNU FP is calculated with the uncorrelated data and is added to the list of PRNU FPs generated from the data set. This leads to a similarity matrix with size $(m - n + 1) \times (m - n + 1)$. This similarity matrix is again clustered using the previously mentioned UCDI, FIC and MCSC clustering techniques.

3.3 Known sensor K-means clustering ((KS)KM)

For this source sensor clustering technique Lloyd's K-means clustering algorithm [27] (KM) has been adopted, as previously proposed in [17]. K-means is a vector quantisation method for cluster analysis used in data mining that partitions n objects into k clusters. The centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimised which leads to a set of clusters that are as compact and well-separated as possible. We define the PRNU noise residuals of the images in the investigated data set as the n objects to cluster, while k is the number of different sensors (clusters). Due to the number of sensors for some data sets is unknown, we repeated the clustering for $k = 1, \ldots, 5$ with the assumption that not more than five sensors have been used. This limitation is not mandatory and can be extended if necessary, but increases the computational effort significantly.

Table 1 List of simulated and existing biometric data sets used in this work with additional information. 'SI' denotes the number of distinct sensor units used for acquiring the images in each data set

Type	Data set name	Number of images	Sensor model	Image size	SI	Sensor type
simulated	*SIMeven*	450	Various Consumer Cameras	$\geq 3264 \times 2448$	3	digital camera
	SIMuneven	450	Various Consumer Cameras	$\geq 3264 \times 2448$	3	digital camera
	SIMdominant	450	Various Consumer Cameras	$\geq 3264 \times 2448$	3	digital camera
existing	*H100_2009*	908	Irisguard H100 IRT	640×480	?	iris sensor
	H100_2013	1451	Irisguard H100 IRT	640×480	1	iris sensor
	IPH_2009	1620	OKI Irispass-h	640×480	?	iris sensor
	IPH_2013	970	OKI Irispass-h	640×480	1	iris sensor
	URU_1	1000	Digital Persona UrU4000 #1	328×356	1	FP sensor
	URU_2	1000	Digital Persona UrU4000 #2	328×356	1	FP sensor

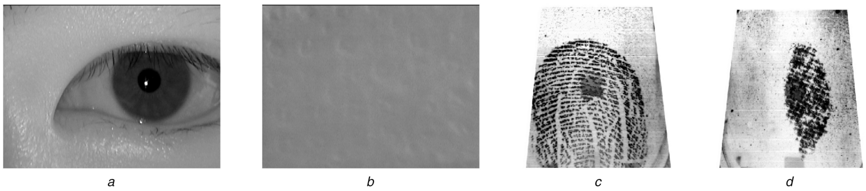


Fig. 2 *Sample images from data sets with additionally acquired uncorrelated data for the corresponding sensor. The 'Digital Persona UrU4000' sensor prevented the acquisition of images without containing at least a partial imprint* ***(a)*** *Image from H100_2009,* ***(b)*** *Uncorrelated data acquired with Irisguard H100 IRT iris sensor,* ***(c)*** *Image from URU_2,* ***(d)*** *Uncorrelated data acquired with Digital Persona UrU4000 #2 FP sensor*

We propose an extension of this technique, the *Known Sensor K-Means Clustering* (KSKM), to be able to make use of the uncorrelated data. We first generate a PRNU FP from the uncorrelated data, which is then added to the set of PRNU noise residuals n which is clustered. In addition, we select this generated PRNU FP as starting point for the algorithm together with $k - 1$ random other samples from the data set. We repeat the K-means algorithm five times with the computed PRNU FP and $k - 1$ randomly chosen samples as starting points to avoid the possibility to get stuck in local minima and the clustering of the best run out of these five is selected as the final result.

4 (Biometric) data sets

First of all, we generated simulated data sets to examine the performance of the source sensor attribution techniques presented in Section 3. These data sets all consist of images from three distinct sensors from a popular Sensor Forensics benchmark database, the Dresden Image Database [28]: Agfa DC-830i, Panasonic DMC-FZ50 and Nikon D200. The data sets all contain randomly selected images from each sensor, where we shuffled chunks of 50 images to obtain a random order. We then generated three different data set types based on the frequency of images from each of the three sensors:

- *SIMeven*: 150 images from each sensor.
- *SIMuneven*: 200 images from the first, 150 from the second and 100 from the third sensor.
- *SIMdominant*: 350 images from one sensor and 50 from the two others each.

We repeated the data set generation ten times for each of the three simulated data set types, where the sensors' order for the image distribution is determined randomly each time, e.g. the sensors providing the most images in the *SIMdominant* data set was chosen randomly each time.

The existing biometric data sets under investigation in this work consist of images for two different biometric modalities, iris and FPs, which are illustrated in Table 1 together with the simulated ones. These biometric data sets have not been published; however, the iris data sets ending with '2013' and FP ones 'URU_1' and 'URU_2' have been acquired during a COST Short-Term Scientific Mission (STSM) as described in [29], while data sets ending with '2009' have been provided by the host institution during the mentioned COST STSM. The ground truth on the number of sensor instances used for the acquisition is only known for the *H100_2013*, *IPH_2013*, *URU_1* and *URU_2* data sets, which consists of one sensor instance. For all other data sets only the sensor model is known, but not how many instances of this model have been used.

All images in this work are 8 bit grey-level JPEG files. The iris data has been collected under near infrared illumination, while the FP sensors used red LEDs. The uncorrelated data used in this work to acquire the PRNU FPs for the known sensors has been acquired according to [29] for the following sensors: *OKI Irispass-h*, *Irisguard H100 IRT*, *Digital Persona UrU4000 #1* and *Digital Persona UrU4000 #2*.

To obtain high-quality PRNU FPs as described by Fridrich [1], images with uncorrelated content and high saturation have been acquired. In some cases the sensor's quality assessment prevented the acquisition of such images, therefore the acquisition was performed in a best effort approach by varying the image content as much as possible to gain a 'cleaner' PRNU FP when averaging the images. Fig. 2 shows exemplary iris and FP images from the existing data sets described above and uncorrelated data acquired with the same sensor. It points out a successful acquisition for the *Irisguard H100 IRT* sensor, and a less successful one for the *Digital Persona UrU4000 #2* sensor.

5 Experimental set-up and results

In the following section, we discuss the results of applying the various source sensor attribution techniques illustrated in Section 3 to the data sets in Section 4. First, we explain the general experimental set-up, which contains a description of the methodology and parameters valid for all experiments. After that we characterise the different experiments conducted in this work, which are divided into two different Sections 5.1 and 5.2.

All the data sets described in Section 4 are investigated independently. The PRNU noise residuals are extracted from a square patch located in the centre of each image. After the extraction the PRNU noise residuals are enhanced using one or more of the techniques mentioned in Section 2. For all clustering techniques where a PRNU FP is generated, in addition PRNU FP enhancements are also applied. The configuration of both enhancement types is described at the beginning of each experiment later on.

For the (KS)BCF and (KS)SWx only clusters containing ten or more images are considered for the final number of clusters results. These techniques are prone to generate a few very small clusters for small PRNU sizes which would have a strong impact on the results because of the overall rather small number of clusters and furthermore, in the investigated biometric scenario, the case that such a small number of images in the data sets is acquired with a different sensor is highly unlikely.

In order to be able to quantitatively assess the clustering of the data sets and reveal differences caused by the various PRNU enhancement techniques the mean silhouette value (MSV) by Rousseeuw [30] has been calculated for each source sensor attribution techniques clustering outcome.

The silhouette value for each point is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters, hence it is a measure between intra- and inter-cluster distances. This technique does not rely on any ground truth information about the clustering of the investigated data set and is therefore well suited for our investigation because the ground truth is not known for all data sets used in this work, which can be seen in Table 1. The result for a single cluster, or $k = 1$, has been determined by calculating the pairwise NCC between all point combinations i and j, where $i \neq j$, and then calculating the mean correlation over all points. For all $k \geq 2$ the MSV for the ith point, S_i, is defined as

$$\mathrm{MSV} = \frac{1}{N} \sum_{n=1}^{N} \frac{b_i - a_i}{\max(a_i, b_i)} \tag{7}$$

where N is the number of noise residuals, a_i is the average distance from the ith point to the other points in the same cluster as i (cohesion), and b_i is the minimum average distance from the ith point to points in a different cluster (separation), minimised over all clusters. The silhouette value ranges from −1 to +1. A high silhouette value indicates that a point i is well-matched to its own cluster, and poorly-matched to neighbouring clusters. If most points have a high silhouette value, then the clustering solution is considered to be an appropriate solution. On the other hand, if many points have a low or negative silhouette value, then the clustering solution may have either too many or too few clusters.

This concludes the general experimental set-up and we will now continue with the discussion of the experimental results for the *Simulated Data Sets*.

5.1 Simulated data sets

The performance evaluation of the source sensor attribution techniques is an important part of this work, since the effects of the advanced PRNU enhancement techniques evaluated later are assessed using the clustering outcome of the different techniques. Hence we applied the various clustering techniques on the simulated data sets *SIMeven*, *SIMuneven* and *SIMdominant*. The PRNU is extracted with the basic $ZM + WF$ configuration, which uses the $F_L uk$ denoising filter, enhances the noise residuals with (ZM) and the PRNU FPs with $ZM + WF$ according to [22].

We measure the performance of the proposed source sensor attribution techniques on the simulated data sets for varying PRNU patches (square size): 64, 128, 256, 512, 768, 1024, 1536 and 2048 pixels. In this case, the resulting scores and the number of clusters are averaged over the ten different randomly generated data sets of each data set type (*SIMeven*, *SIMuneven* and *SIMdominant*) separately.

For the simulated data sets, where the ground truth on the source sensor for each image is known, we compute the V-measure (*VM*) [31] score for the clustering outcome, which is defined as harmonic mean of homogeneity (h) and completeness (c) as shown in the following equation:

$$VM = 2 * \frac{h * c}{h + c}, \quad h = 1 - \frac{H(C|K)}{H(C)}, \quad c = 1 - \frac{H(K|C)}{H(K)} \tag{8}$$

The homogeneity h measures whether each cluster exclusively contains images from the same sensor, while the completeness c measures if all images belonging a sensor have been assigned to the same cluster. $H(C|K)$ refers to the conditional entropy of the different classes for the given cluster associations and $H(C)$ denotes the entropy of the classes. Further details can be found in the corresponding paper [31].

First of all we have a look at how the size of the extracted PRNU affects the performance. Since the simulated data sets contain higher resolution images than the biometric data we are able to test various extracted PRNU sizes from 64×64 to 2048×2048 pixels. The results show that the VM scores increase proportionally with the PRNU size for some techniques, where BCF shows a steady increase in clustering performance with increasing PRNU size, while for KM the performance increases until a certain point and then stagnates. The stagnation of the VM scores after a certain PRNU size occurs due to the technique's inability to further exploit the additional data for the differentiation of the sensors in the data. Thus it reaches a point where additional data does not change the cluster association of the images.

The MSV scores in general increase with larger PRNU size, except for the KM technique. The decreasing MSV scores for the KM technique with larger PRNU sizes can be explained by how the MSV scores are calculated. For the MSV scores we consider pairwise Euclidean distances between the PRNU noise residuals, which become more and more inaccurate with increasing dimensionality (i.e. PRNU size), as shown in [32]. Due to the cluster association staying the same for larger PRNU sizes, the MSV scores decrease because of this effect in higher dimensions. For the SWx techniques the MSV score increases with higher dimension because of their inability to cluster the data properly.

For the SWx techniques, the VM performance is consistently bad across all tested PRNU sizes. The reason for this are the very low homogeneity scores for the SWU, the very low completeness scores for SWF and while SWM shows the best VM score of the three, but suffers from both mediocre homogeneity and completeness scores. The VM and MSV results for BCF, KM and SWF are illustrated in Fig. 3.

Due to the limit of the biometric data to extract the PRNU from a 256×256 patch we compared the performance of all techniques with this configuration, which can be seen in Fig. 4. It shows that the highest VM score is obtained by the KM technique, which shows a high score for the *SIMeven* and *SIMuneven* data sets, while it seems to struggle with the *SIMdominant* data set. In general, all techniques obtain much lower scores for the *SIMdominant* data set with BCF being the only exception. Although the SWU and SWM generate a number of clusters close to the expected result of 3, the quality of the clusters in respect to the homogeneity and completeness is quite low. BCF on the other hand generates a few more clusters, but their quality is higher, which is indicated by the higher VM score.

Summarising the KM and BCF techniques are the most qualified techniques to cluster the data for the tested PRNU size. The KM technique obtains the highest scores for all three simulated data sets, but the performance varies highly depending on the distribution of the images from different sensors within the data sets. The BCF technique on the other hand performs worse than the KM one due to being prone to produce more clusters, which is penalised by the VM measure. However, the produced clusters all have a high homogeneity and by having the most consistent results across all the simulated data sets still consider this method as well suited for the clustering. Due to the poor results for the SWx techniques they cannot be recommended for this kind of scenario, thus for the remaining evaluation only the BCF and KM techniques are taken into consideration.

5.2 Iris and FP data sets

In this section, *Iris and FP Data Sets*, we discuss the effects of applying different PRNU enhancement techniques on the existing biometric data sets. For these iris and FP data sets we are only able to extract 256×256 pixel patches because of the varying image size to ensure the comparability of the results among all data sets.

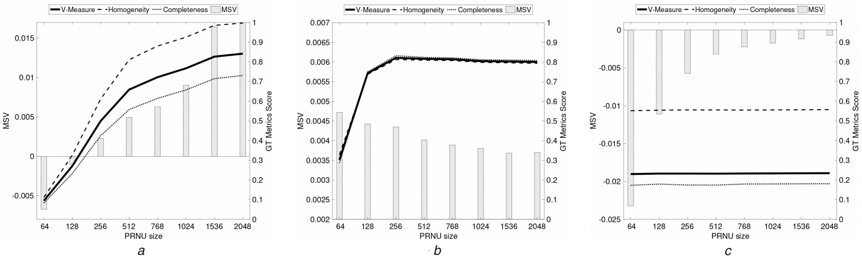


Fig. 3 *Averaged MSV and computed ground truth (GT) metrics (V – measure, homogeneity, completeness) examples for three different source sensor attribution techniques applied to the simulated data sets SIMeven, SIMuneven and SIMdominant using different sizes for the extracted PRNU and the basic ZM + WF PRNU enhancement configuration. The scores have been averaged over the three data sets* ***(a)*** *BCF,* ***(b)*** *KM,* ***(c)*** *SWF*

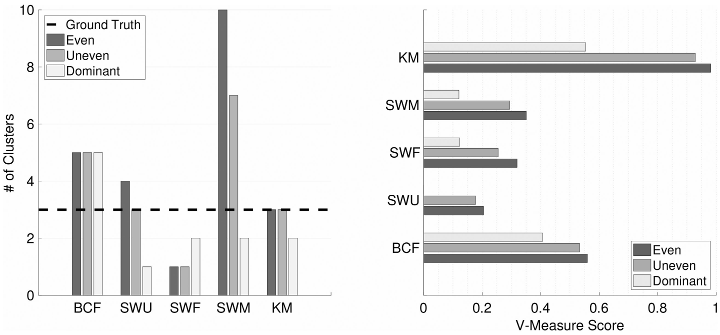


Fig. 4 *Resulting number of clusters (right) and V-measure scores (left) for each of the source sensor attribution techniques applied on the simulated data sets using a* 256 × 256 *pixel PRNU patch size and the basic ZM + WF PRNU enhancement configuration. The expected number of sensors used to create the data sets (3) is shown as the dashed line (ground truth)*

Table 2 Enhancement configurations applied for the different steps of the PRNU extraction process. The abbreviations are explained in Section 2

Name	Denoising filter	Noise residuals	FPs
ZM + WF [22]	F_{Luk}	ZM	WF
Li [7]	F_{Luk}	Li	—
BM3D [6]	F_{BM3D}	—	—
FSTV [9]	F_{FSTV}	—	—
FDR + SEA [12]	F_{Mih}	FDR + Li	SEA

The different configurations for the PRNU extraction process used for the experiments can be seen in Table 2. The parameters of all PRNU enhancement techniques have been chosen as recommended by the authors of the respective papers.

This section is further divided into the following three subsections:

- In Section 5.2.1, we briefly evaluate the results obtained with the basic *ZM + WF* configuration applied for the PRNU extraction for all clustering techniques.
- Section 5.2.2 discusses the effects of the different PRNU extraction configurations applied for all clustering techniques.
- In Section 5.2.3, we recapitulate the effects of the various PRNU extraction configurations and compare their performance across all data sets.

Before discussing the *Baseline* results, an overview overall results for the biometric data sets is given in Table 3, where we will depict some interesting observations in the following.

5.2.1 Baseline: The resulting MSV values relevant for the *Baseline* evaluation correspond to the *ZM + WF* rows of Table 3. The resulting clusters for all source sensor attribution techniques can be seen in Fig. 5.

First of all we have a look at the iris and FP data set results separately. The first thing we notice when looking at the iris data sets is that the BCF and KSBCF techniques produce a large number of clusters for the *IPH_2009* and *IPH_2013*, where both are not able to cluster the data properly. This is also confirmed by the negative MSV scores. However, the use of uncorrelated data helps to improve the MSV scores slightly for KSBCF compared to BCF. KM and KSKM yield one cluster for all iris data sets, even for those with known ground truth that have been acquired with a single sensor. The use of uncorrelated data does not affect the MSV scores at all for the KSKM technique compared to KM.

For the FP data sets *URU_1* and *URU_2* all clustering techniques fail at clustering the data correctly and yield two clusters, even though the correct number would be 1. Yet all MSV scores are positive which indicates that the separation of the data into two clusters could be reasonable. The effects of the uncorrelated data are the same as for the iris data sets, where the MSV scores of KSBCF are slightly better than those for BCF and the MSV scores for KSKM do not show any change in comparison with KM.

5.2.2 PRNU enhancements side by side: In this subsection, we will have a look at the *Li*, *BM3D*, *FSTV* and *FDR + SEA* rows of Table 3, which contain the results of applying the PRNU extraction configurations described in Table 2. The evaluation of the results focuses on the BCF and KSBCF techniques first, followed by the KM and KSKM techniques.

The results for the BCF and KSBCF techniques are graphically depicted in Fig. 6. As we can see the BCF results for *H100_2009*

Table 3 MSVs for the various combinations of PRNU enhancement configurations and source sensor attribution techniques. The numbers in parentheses show the ground truth number of clusters in the table header and the number of clusters generated for the different combinations of source sensor attribution techniques and data sets in the table body

		H100_2009 (?)	H100_2013 (1)	IPH_2009 (?)	IPH_2013 (1)	URU_1 (1)	URU_2 (1)
BCF	ZM + WF	0.0161 (1)	0.0122 (1)	−0.0020 (6)	−0.0025 (6)	0.0035 (2)	0.0077 (2)
	Li	0.0171 (1)	0.0121 (1)	0.0024 (2)	0.0026 (2)	0.0065 (1)	0.0057 (2)
	BM3D	0.0197 (1)	0.0160 (1)	−0.0012 (2)	0.0011 (2)	0.0149 (1)	0.0115 (2)
	FSTV	−0.0401 (2)	−0.0004 (2)	−0.0002 (2)	0.0030 (2)	0.0095 (1)	0.0086 (2)
	FDR + SEA	0.0169 (1)	0.0120 (1)	−0.0009 (5)	0.0001 (4)	0.0139 (1)	0.0076 (2)
KSBCF	ZM + WF	0.0161 (1)	0.0122 (1)	−0.0015 (6)	−0.0019 (4)	0.0090 (2)	0.0088 (2)
	Li	0.0171 (1)	0.0121 (1)	0.0042 (1)	0.0095 (1)	0.0045 (2)	0.0051 (2)
	BM3D	0.0197 (1)	0.0160 (1)	0.0016 (2)	0.0023 (1)	0.0068 (2)	0.0101 (2)
	FSTV	−0.0660 (1)	0.0377 (1)	0.0019 (1)	0.0015 (1)	0.0044 (2)	0.0077 (2)
	FDR + SEA	0.0169 (1)	0.0120 (1)	0.0019 (5)	0.0008 (3)	0.0060 (2)	0.0066 (2)
KM	ZM	0.0161 (1)	0.0122 (1)	0.0036 (1)	0.0035 (1)	0.0291 (2)	0.0213 (2)
	Li	0.0171 (1)	0.0121 (1)	0.0207 (1)	0.0215 (1)	0.0242 (2)	0.0177 (2)
	BM3D	0.0197 (1)	0.0160 (1)	0.0214 (1)	0.0228 (1)	0.0414 (2)	0.0275 (2)
	FSTV	0.0344 (2)	0.0187 (1)	0.0230 (1)	0.0245 (1)	0.0347 (2)	0.0255 (2)
	FDR	0.0169 (1)	0.0120 (1)	0.0222 (1)	0.0233 (1)	0.0355 (2)	0.0259 (2)
KSKM	ZM + WF	0.0161 (1)	0.0122 (1)	0.0036 (1)	0.0035 (1)	0.0291 (2)	0.0213 (2)
	Li	0.0171 (1)	0.0121 (1)	0.0207 (1)	0.0215 (1)	0.0242 (2)	0.0177 (2)
	BM3D	0.0197 (1)	0.0160 (1)	0.0214 (1)	0.0228 (1)	0.0414 (2)	0.0275 (2)
	FSTV	0.0344 (2)	0.0187 (1)	0.0230 (1)	0.0245 (1)	0.0347 (2)	0.0255 (2)
	FDR + SEA	0.0169 (1)	0.0120 (1)	0.0222 (1)	0.0233 (1)	0.0355 (2)	0.0259 (2)

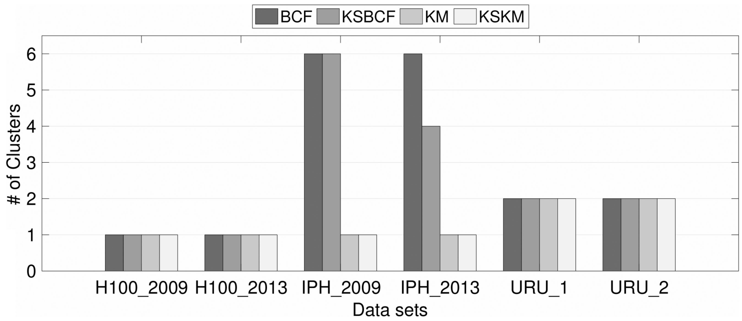
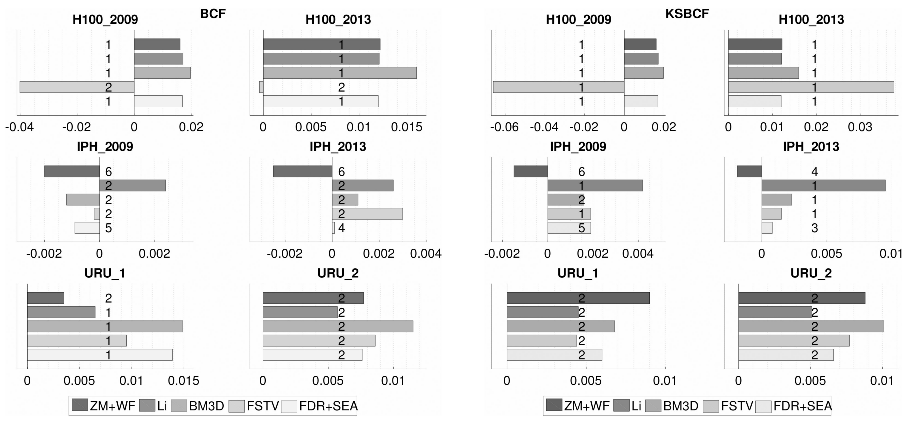


Fig. 5 *Clustering result with number of clusters for the Baseline evaluation (ZM + WF)*

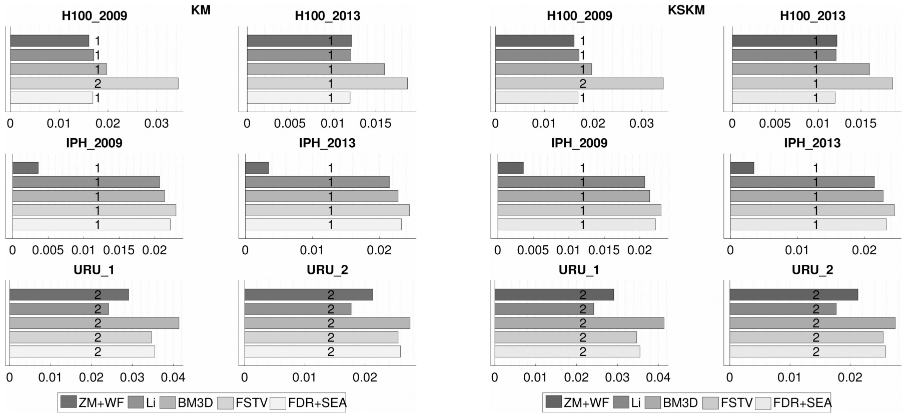
and *H100_2013* are quite similar, where *BM3D* obtains the highest MSV scores. However, all the other configurations are quite close with exception of *FSTV*, which yields an increased number of clusters 2 accompanied by negative MSV scores. The use of uncorrelated data in KSBCF has no effect on the MSV scores nor on the number of clusters for all configurations except FSTV, where it reduces the number of clusters to 1 in both data sets. The MSV scores for KSBCF and the FSTV configuration; however, are even further reduced for the *H100_2009*, while it shows a dramatic increase for the *H100_2013* data set where it even surpasses the previously best score of *BM3D* by a clear margin. The *IPH_2009* and *IPH_2013* have been very challenging for the basic *ZM + WF* and showed poor results due to the high number of clusters. All applied PRNU enhancement configurations show a MSV score improvement with all configurations for the BCF technique by decreasing the number of clusters even down to 2 in most cases. These results can be explained by the improved separability of the data due to the PRNU enhancements. The best MSV score for *IPH_2009* is obtained by *Li*, while for *IPH_2013 FSTV* shows the highest MSV with the *Li* configuration being close. With the addition of uncorrelated data in KSBCF the number of clusters is decreased to 1 for some configurations in both data sets, which in the case of *IPH_2013* corresponds to the ground truth. *Li* takes the most advantage of these data and yields the by far highest MSV scores for both the *IPH_2009* and *IPH_2013* data sets. For the last two data sets, *URU_1* and *URU_2*, BCF is able to improve the MSV scores for almost all PRNU enhancement configurations. The number of clusters is decreased to 1 for *URU_1*; however, for *URU_2* the number stays at 2, which is incorrect according to the ground truth. The highest MSV scores for both scores are obtained with the *BM3D* configuration. Interestingly, the addition of uncorrelated data in KSBCF lowers the scores for all PRNU enhancement configurations compared to *ZM + WF*. Due to the number of clusters remains constant, the differentiability of the data could be negatively affected by the suboptimal capturing of the uncorrelated data for the FP sensors, as shown in Fig. 2. *ZM + WF* yields the highest MSV score for *URU_1* and *BM3D* for *URU_2*.

The second part of the evaluation looks at the results of the KM and KSKM clustering techniques, which are illustrated in Fig. 7. The first thing that we notice here is that the use of uncorrelated data in the KSKM technique has absolutely no effect on the scores and the number of clusters. Therefore, all of the following statements relate to both KM and KSKM. In most cases, the PRNU enhancement configurations show an improvement of the MSV scores, while not changing the resulting number of clusters. The only exception is *FSTV*, which increases the number of clusters to 2 for the *H100_2009* data set. The highest MSV scores for the iris data sets (*H100_2009*, *H100_2013*, *IPH_2009* and *IPH_2013*) are



a *b*

Fig. 6 *Comparison of number of clusters (at the centre of each bar chart) and MSV values for the Li, BM3D, FSTV and FDR + SEA PNRU enhancement configurations compared to ZM + WF when applied in the BCF and KSBCF techniques **(a)** BCF, **(b)** KSBCF*



a *b*

Fig. 7 *Comparison of number of clusters (at the centre of each bar chart) and MSV values for the Li, BM3D, FSTV and FDR + SEA PNRU enhancement configurations compared to ZM + WF when applied in the KM and KSKM techniques **(a)** KM, **(b)** KSKM*

obtained by *BM3D* and by *FSTV* for the FP data sets (*URU_1* and *URU_2*).

5.2.3 Summary biometric data: The preceding results for the biometric show that the adoption of the different PRNU enhancement configurations did indeed help to improve the clustering outcome of the clustering techniques. Fig. 8 shows the PRNU enhancement configurations resulting in the highest MSV scores for each technique and data set. We can see that for the KM and KSKM technique *FSTV* is the best choice for the iris and *BM3D* for the FP data sets. Regarding the BCF and KSBCF the choice of PRNU enhancement configuration is dependent on the data set or rather on the sensor model: For the data sets using the *Irisguard H100 IRT* sensor *BM3D* is the configuration of choice, while for the *OKI Irispass-h* sensor it is the *Li* configuration.

The additional use of uncorrelated data had a very large impact on the clustering outcome of the KSBCF techniques applied on the *IPH_2009* and *IPH_2013* data sets compared to BCF. However, for the other data sets the impact was quite small and for KSKM the uncorrelated data had no impact at all. This can be explained by how the KSKM technique makes use of the uncorrelated data, in fact, it is only used to create a starting point for the K-means algorithm which then nevertheless converges to the same cluster centroids as without using this additional data.

Concerning the data sets for which the ground truth is known, the correct number of clusters for all iris data sets could be determined at least by applying a combination of uncorrelated data and PRNU enhancements. In contrast, for the FP data set the correct number could only be established in one case. In all the others the clustering techniques failed to do so even with any combination of uncorrelated data and PRNU enhancement.

Recapitulating we can say that there is no single best PRNU enhancement configuration for this scenario, yet it is highly situational which one should be chosen.

6 Conclusion

In this work, we proposed novel source sensor attribution techniques based on the sensors PRNU and applied existing ones. We generated multiple simulated data sets containing images from multiple sensors taken from the Dresden Image Database and computed different clustering quality metrics to evaluate the

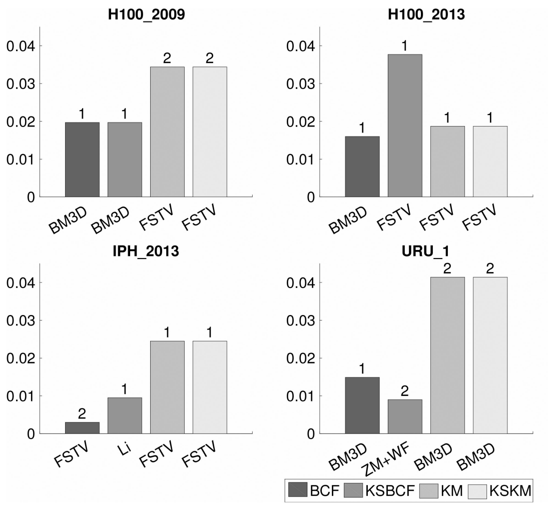


Fig. 8 *Highest MSV scores, according number of clusters (above bars) and PRNU enhancement configuration achieving the score for all clustering technique and data set combinations*

proposed techniques. The results showed that the size of the extracted PRNU has a significant impact on the clustering result. Two of the techniques BCF and KM have been able to cluster the data properly and showed consistent and promising results in the case of 256 × 256 PRNU patch sizes and have been considered appropriate for the source sensor attribution of biometric sensors.

In the following, all techniques have been applied to biometric data sets with low resolution images of two different biometric modalities, iris and FPs, to cluster the images according to their source sensor. Different PRNU enhancement techniques have been adopted in response to the special characteristics of biometric data, such as highly correlated data and contamination of the PRNU by the image content, in order to improve the clustering performance. In addition, we used uncorrelated data acquired with the sensors and proposed several extensions for already existing sensor attribution techniques to be able to use this uncorrelated data in conjunction with the source attribution techniques.

The evaluation of the various PRNU enhancement and uncorrelated data effects was conducted by means of a quantitative measure for the clustering outcome that considers the cohesion and separation of the clusters without the need of any knowledge about the underlying cluster ground truth. Summarising the results it can be stated that most PRNU enhancements did indeed help to improve the clustering results compared to the original work in [21] by increasing the differentiability of the PRNU noise residuals. However, we could not identify any single enhancement technique or combination that was able to improve the clustering outcome for all data sets alike, but the choice of the best performing technique is highly situational. Furthermore, the clustering techniques in most cases did not succeed in determining the correct number of clusters for the FP data sets, even with the support of the different PRNU enhancements techniques.

For the FP data sets the absent PRNU enhancement effect and poor results clearly needs some further and deeper investigation. The insufficient quality of the extracted PRNU might be an issue in this case, either caused by the image content or by other contaminations or factors, e.g. the amount of denoising applied during the biometric sensor's processing of the acquired image. Since biometric sensors are often closed systems tailored to acquire a specific type of images, the identification of these issues is challenging. In conclusion certainly further studies have to be conducted in this manner in regard to the special requirements posed by biometric sensors and the data they produce. A fusion of the source sensor attribution techniques' clustering outcome will also be investigated in future work.

7 Acknowledgment

This work was partially funded by the Austrian Science Fund (FWF) under project no. P26630 and partially supported by a COST 1106 Short-Term Scientific Mission (STSM).

8 References

[1] Fridrich, J.: 'Sensor defects in digital image forensics', in Sencar, H.T., Memon, N. (Eds.): '*Digital image forensics: there is more to a picture than meets the eye*' (Springer Verlag, 2012), ch. 6, pp. 179–218

[2] Bartlow, N., Kalka, N., Cukic, B., *et al.*: 'Identifying sensors from fingerprint images'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, 2009, CVPR Workshops 2009, June 2009, pp. 78–84

[3] Uhl, A., Höller, Y.: 'Iris-sensor authentication using camera PRNU fingerprints'. Proc. of the 5th IAPR/IEEE Int. Conf. on Biometrics (ICB'12), New Delhi, India, March 2012, pp. 1–8

[4] Kalka, N., Bartlow, N., Cukic, B., *et al.*: 'A preliminary study on identifying sensors from iris images'. The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015

[5] Gloe, T., Pfennig, S., Kirchner, M.: 'Unexpected artefacts in prnu-based camera identification: a 'Dresden image database' case-study'. MM&Sec'12: Proc. of the 14th ACM Multimedia and Security Workshop, September 2012, pp. 109–114

[6] Dabov, K., Foi, A., Katkovnik, V., *et al.*: 'Image denoising by sparse 3-d transform domain collaborative filtering', *IEEE Trans. Image Process.*, 2007, **16**, (8), pp. 2080–2095

[7] Li, Ch.-T.: 'Source camera identification using enhanced sensor pattern noise', *IEEE Trans. Inf. Forensics Sec.*, 2010, **5**, (2), pp. 280–287

[8] Cooper, A.J.: 'Improved photo response non-uniformity (prnu) based source camera identification', *Forensic Sci. Int.*, 2013, **226**, (13), pp. 132–141

[9] Gisolf, F., Malgoezar, A., Baar, T., *et al.*: 'Improving source camera identification using a simplified total variation based noise removal algorithm', *Digital Invest.*, 2013, **10**, (3), pp. 207–214

[10] Kang, X., Chen, J., Lin, K., *et al.*: 'A context-adaptive spn predictor for trustworthy source camera identification', *EURASIP J. Image Video Process.*, 2014, **2014**, (1), p. 19

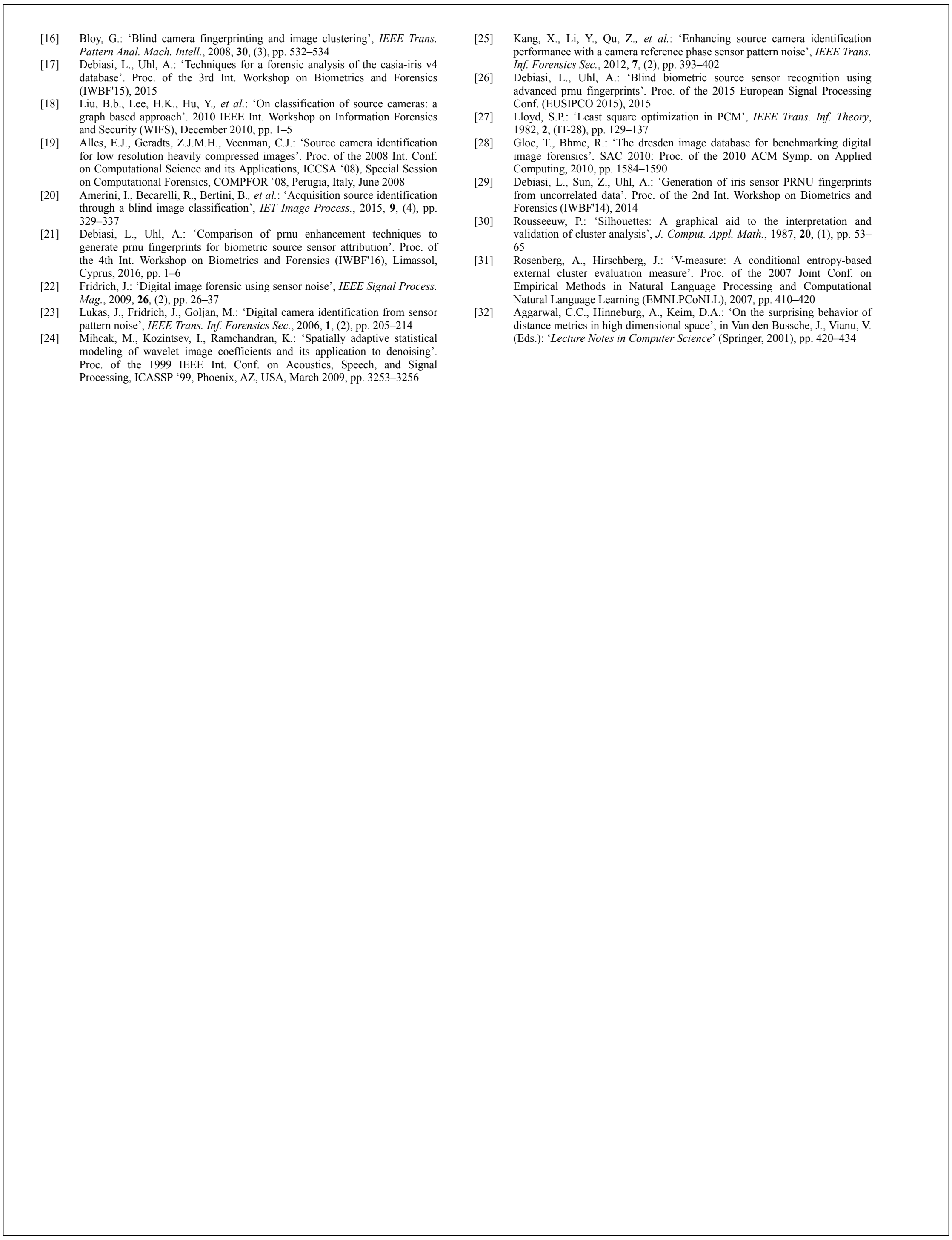
[11] Lin, X., Li, Ch.-T.: 'Preprocessing reference sensor pattern noise via spectrum equalization', *IEEE Trans. Inf. Forensics Sec.*, 2016, **11**, (1), pp. 126–140

[12] Lin, X., Li, Ch.-T.: 'Enhancing sensor pattern noise via filtering distortion removal', *IEEE Signal Process. Lett.*, 2016, **23**, (3), pp. 381–385

[13] Amerini, I., Caldelli, R., Crescenzi, P., *et al.*: 'Blind image clustering based on the normalized cuts criterion for camera identification', *Signal Process. Image Commun.*, 2014, **29**, (8), pp. 831–843

[14] Li, Ch.-T.: 'Unsupervised classification of digital images using enhanced sensor pattern noise'. ISCAS, 2010, pp. 3429–3432

[15] Caldelli, R., Amerini, I., Picchioni, F., *et al.*: 'Fast image clustering of unknown source images'. IEEE Int. Workshop on Information Forensics and Security (WIFS) 2010, 2010, pp. 1–5



[16] Bloy, G.: ‘Blind camera fingerprinting and image clustering’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (3), pp. 532–534

[17] Debiasi, L., Uhl, A.: ‘Techniques for a forensic analysis of the casia-iris v4 database’. Proc. of the 3rd Int. Workshop on Biometrics and Forensics (IWBF'15), 2015

[18] Liu, B.b., Lee, H.K., Hu, Y., *et al.*: ‘On classification of source cameras: a graph based approach’. 2010 IEEE Int. Workshop on Information Forensics and Security (WIFS), December 2010, pp. 1–5

[19] Alles, E.J., Geradts, Z.J.M.H., Veenman, C.J.: ‘Source camera identification for low resolution heavily compressed images’. Proc. of the 2008 Int. Conf. on Computational Science and its Applications, ICCSA ‘08), Special Session on Computational Forensics, COMPFOR ‘08, Perugia, Italy, June 2008

[20] Amerini, I., Becarelli, R., Bertini, B., *et al.*: ‘Acquisition source identification through a blind image classification’, *IET Image Process.*, 2015, **9**, (4), pp. 329–337

[21] Debiasi, L., Uhl, A.: ‘Comparison of prnu enhancement techniques to generate prnu fingerprints for biometric source sensor attribution’. Proc. of the 4th Int. Workshop on Biometrics and Forensics (IWBF'16), Limassol, Cyprus, 2016, pp. 1–6

[22] Fridrich, J.: ‘Digital image forensic using sensor noise’, *IEEE Signal Process. Mag.*, 2009, **26**, (2), pp. 26–37

[23] Lukas, J., Fridrich, J., Goljan, M.: ‘Digital camera identification from sensor pattern noise’, *IEEE Trans. Inf. Forensics Sec.*, 2006, **1**, (2), pp. 205–214

[24] Mihcak, M., Kozintsev, I., Ramchandran, K.: ‘Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising’. Proc. of the 1999 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP ‘99, Phoenix, AZ, USA, March 2009, pp. 3253–3256

[25] Kang, X., Li, Y., Qu, Z., *et al.*: ‘Enhancing source camera identification performance with a camera reference phase sensor pattern noise’, *IEEE Trans. Inf. Forensics Sec.*, 2012, **7**, (2), pp. 393–402

[26] Debiasi, L., Uhl, A.: ‘Blind biometric source sensor recognition using advanced prnu fingerprints’. Proc. of the 2015 European Signal Processing Conf. (EUSIPCO 2015), 2015

[27] Lloyd, S.P.: ‘Least square optimization in PCM’, *IEEE Trans. Inf. Theory*, 1982, **2**, (IT-28), pp. 129–137

[28] Gloe, T., Bhme, R.: ‘The dresden image database for benchmarking digital image forensics’. SAC 2010: Proc. of the 2010 ACM Symp. on Applied Computing, 2010, pp. 1584–1590

[29] Debiasi, L., Sun, Z., Uhl, A.: ‘Generation of iris sensor PRNU fingerprints from uncorrelated data’. Proc. of the 2nd Int. Workshop on Biometrics and Forensics (IWBF'14), 2014

[30] Rousseeuw, P.: ‘Silhouettes: A graphical aid to the interpretation and validation of cluster analysis’, *J. Comput. Appl. Math.*, 1987, **20**, (1), pp. 53–65

[31] Rosenberg, A., Hirschberg, J.: ‘V-measure: A conditional entropy-based external cluster evaluation measure’. Proc. of the 2007 Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLPCoNLL), 2007, pp. 410–420

[32] Aggarwal, C.C., Hinneburg, A., Keim, D.A.: ‘On the surprising behavior of distance metrics in high dimensional space’, in Van den Bussche, J., Vianu, V. (Eds.): ‘*Lecture Notes in Computer Science*’ (Springer, 2001), pp. 420–434

Chapter 21

Identifying Iris Sensors from Iris Images

Luca Debiasi, Christof Kauba and Andreas Uhl

The base component of iris sensors deployed in practical applications is a digital image sensor, mostly supported by a near infra-red (NIR) light source to improve the iris recognition results [1]. These sensors acquire digital images, which are then further processed and inserted into a biometric system's processing chain.

The authenticity and integrity of the acquired iris images plays an important role for the overall security of a biometric system. Ratha *et al.* [2] identified eight stages in a generic biometric system where attacks may occur. Figure 21.1 shows an insertion and presentation attack on an exemplary biometric system. An insertion attack bypasses the biometric sensor by inserting data (biometric sample) into the transmission from the sensor to the feature extractor. This transmission is the most relevant point for an attack on the integrity and authenticity of the acquired iris images, where the iris image inserted during the attack could be acquired with another sensor off-site, even without the knowledge of a genuine user, or a manipulated image to spoof the biometric recognition system. In contrast to the insertion attack, in case of the presentation attack a forged or fake biometric trait, i.e. an artificially manufactured fake fingerprint or a print of an iris image, is presented to the genuine sensor installed in the biometric system. The presentation of a forged biometric trait can usually be detected by deploying different liveness detection systems.

Encryption and other classical authentication techniques like digital signatures or data-hiding have been suggested to secure the previously mentioned transmis-

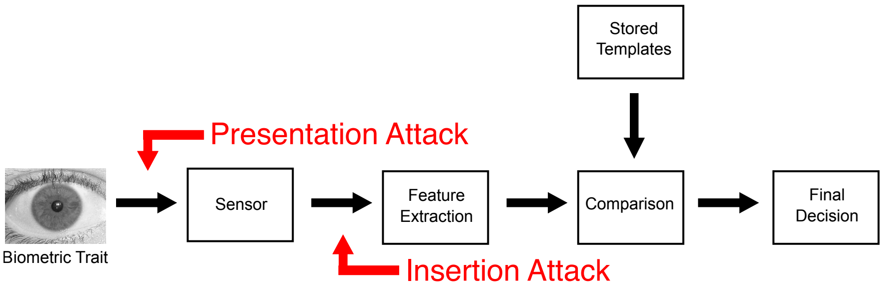


Figure 21.1: Exemplary biometric system and point of insertion and presentation attacks.

sion channel by verifying the senders (i.e. sensor and feature extractor) authenticity, as well as the integrity of the entire authentication mechanism. The proposed approaches can be divided into active and passive-blind approaches.

Active methods consist of data hiding approaches [3, 4] and the digital signature approaches [5, 6, 7, 8]. Höller *et al.* [9] describe the pros and cons of these active methods as follows:

- **Classical digital signatures** work by adding additional data to verify the original data, whereas watermarks become an integral part of the sample data, and moreover, spatial locations of eventual tampering can be identified [10].
- **Fragile watermarks** (as proposed for these tasks in e.g. [11, 12, 13]) cannot provide any form of robustness against channel errors and unintentional signal processing "attacks" like compression, which is the same as with classical digital signatures.
- **Semi-fragile watermarks** have been designed to differentiate between allowed signal processing operations and malicious attacks and have also been suggested for employment in biometric systems [14, 15, 16, 17].

Höller *et al.* [9] also mention that a general drawback of watermarks is the representation of additional data which is inserted into the sample data, where an impact on recognition accuracy may be expected. In fact, literature reports on corresponding effects in case of iris recognition [18], speech recognition [19], and fingerprint recognition [20].

Passive-blind approaches, in contrast to active methods, do not need any prior information about the image. As stated in [21], passive-blind approaches are mostly based on the fact that forgeries can bring specific detectable changes into the image (e.g., statistical changes). In high quality forgeries, these changes cannot be found by visual inspection.

The field of digital image forensics deals with still images and analyzing traces in still image data. Two major tasks in this field are establishing an image's origin and its integrity. In contrast to digital watermarking as authenticity technique, as mentioned in [22], digital image forensics do not require any active embedding step at the time of creation or publication. Evidence is extracted merely from structural analysis of image files and statistical analysis of the image data (i. e. the two-dimensional array of pixel intensities).

To determine an image's origin several approaches have been proposed exploiting hardware and software related artifacts. Investigated hardware related artifacts cover optical defects, like chromatic aberrations [23] or lens distortions [24], or sensor artifacts, like sensor defects [25] and noise. Software artifacts are introduced during the processing of the images in the cameras and can be unveiled using statistical features [26] or by analysing the common image processing pipeline of the images [27].

The photo-response non-uniformity (PRNU) of imaging sensors, as described in [28, 29], is an intrinsic property of all digital imaging sensors due to slight variations among individual pixels in their ability to convert photons to electrons. Consequently, every sensor casts a weak noise-like pattern onto every image it takes.

This pattern, which plays the role of a “sensor fingerprint”, is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering. This fingerprint can be estimated from images taken by the camera and later detected in a given image to establish image origin and integrity.

Even though the PRNU is stochastic in nature, it is a relatively stable component of the sensor over its life span, providing a unique sensor fingerprint with the following important properties [29]:

1. **Dimensionality**: The fingerprint is stochastic in nature and has a large information content, which makes it unique to each sensor.
2. **Universality**: All imaging sensors exhibit PRNU.
3. **Generality**: The fingerprint is present in every picture independently of the camera optics, camera settings, or scene content, with the exception of completely dark images.
4. **Stability**: It is stable in time (except for ageing related sensor defects) and under a wide range of environmental conditions (temperature, humidity, etc.).
5. **Robustness**: It survives lossy compression, filtering, gamma correction, and many other typical processing procedures.

Slight variations of individual pixels during the conversion of photons to electrons in digital image sensors are the source of the PRNU, thus it is considered an intrinsic property which is present in all digital imaging sensors. Every digital image sensor adds this weak, noise-like pattern into every image acquired with it. The sensor identification can be performed at different levels, as described by Bartlow *et al.* [30]: Technology, brand, model, unit. Due to the datasets evaluated in this work we focus on the model level, which corresponds to a differentiation according to model and brand.

The PRNU can also be used for the verification of an image’s integrity. The integrity is compromised if an image has been geometrically transformed (e.g. cropped, rotated, turned, flipped etc.) or if parts of the image have been tampered (e.g. deleted, copied, replaced, altered). These manipulations lead to changes in the PRNU which can be detected as shown in [31, 32].

In the context of biometric systems security the PRNU fingerprint of a sensor can be used to ensure the integrity and authenticity of images acquired with a biometric sensor. Höller *et al.* [9] propose a suitable passive approach to secure the transmission channel between the sensor and the feature extractor, making use of sensor fingerprints based on a sensor’s PRNU [31]. Besides image integrity, this technique can also provide authenticity by identifying the source sensor uniquely and important properties as required in a biometric scenario have been demonstrated: suitability to manage large datasets [33, 34], robustness against common signal processing operations like compression and malicious signal processing [35, 36], and finally methodology to reveal forged PRNU fingerprints has been established [37].

To ensure the authenticity of the biometric sensor, first the discriminative power of the biometric sensors has to be evaluated, as it has been done in [9] and [30] using the PRNU. The results from Höller *et al.* [9], where the discriminative power of five iris sensors from the *CASIA-Iris V4* database has been evaluated show high variations. Other work by Kalka *et al.* [38] regarding the differentiability of iris sensor showed varying results, while studies conducted on fingerprint sensors by Bartlow *et al.* [30] showed more satisfactory results. In order for PRNU fingerprints beeing useful as an authentication measure for biometric systems, the sources of the poor differentiation results have to be determined. Some possible explanations are given in [38] and [9] and consist of the highly correlated data of biometric datasets, saturated pixels and the use of multiple sensors of the same model. An additional caveat for the PRNU extraction is the image content. Since the PRNU covers the high frequency components of an image, it is contaminated with other high frequency components within the images, such as edges. Li [39] proposed an approach for attenuating the influence of details from scenes on the PRNU so as to improve the device identification rate of the identifier. Moreover the PRNU fingerprint can be extracted from images of a biometric sensor and injected into forged images, as described by Goljan *et al.* [40]. Using several images captured by the sensor deployed in the biometric system a suitable PRNU fingerprint can be generated by the attacker. This attack can only be detected with a triangle test [9], which requires additional genuine images acquired under controlled conditions.

To overcome the reported problems with the PRNU extraction for some sensors and the injection attack, there exist other approaches like the one by El-Naggar and Ross [41], who proposed a passive approach tailored to iris recognition. At first the ocular image is segmented to get the iris region, then the iris texture is unwrapped, followed by a normalisation step to get a normalised iris image. Only the inner half of this normalised iris image is used and further split into a set of overlapping blocks. For each block 50 Gabor and 68 statistical features are extracted to form a 118 dimensional feature vector representing the iris image. These feature vectors are then classified using a 3-layer artificial neural network. They were able to achieve accuracies of $80-85\%$. We propose a similar approach which follows their evaluation methodology but uses different features and a SVM classifier. In this chapter we evaluate this approach and a PRNU based one, trying to identify the iris data set which an iris image belongs to. If the correct data set can be determined for a given iris image, these approaches could be used to secure an iris recognition system against insertion attacks.

The chapter is organised as follows: In Section 21.1 we describe the two different approaches and the examined iris data sets are listed in Section 21.2. The experimental setup and the results are illustrated in Section 21.3 and 21.4, respectively. In Section 21.5 we discuss how the previously examined techniques can be used in a practical application. Finally Section 21.6 concludes the chapter.

21.1 Techniques for Sensor Identification/Dataset Classification

In this section we present two different techniques that allow to infer which dataset an iris image originates from. The first technique, PRNU based Sensor Identification (PSI), does this by identifying the sensor used to acquire the image. The second technique, Iris Texture Classification (ITC), makes use of the iris texture and its inherent features to classify the iris images according to the source sensor. Both techniques are presented in detail in the following section.

21.1.1 PRNU based Sensor Identification (PSI)

A digital image sensor consists of lots of small photosensitive, usually rectangular detectors that capture the incident light and generate an electric signal. These detectors are commonly known as pixels. The image acquired with the sensor is constructed by the aggregate of all pixels. Due to imperfections in the manufacturing and the inhomogeneity of the manufacturing material, silicon, the efficiency of each pixel of converting photons to electrons varies slightly. According to Fridrich [29], the raw output of a sensor with $w \times h$ pixels can be modeled as:

$$\begin{aligned} & Y = I + I \circ K + \tau D + C + \Theta \\ with \quad & Y, I, K, D, C, \Theta \in \mathbb{R}^{w \times h}; \tau \in \mathbb{R} \end{aligned} \tag{21.1}$$

where Y is the sensor output (image). I represents the incoming light, $I \circ K$ the photo-response non-uniformity PRNU, τD the dark current (with τ being a multiplicative factor representing exposure settings, sensor temperature, etc.). The matrix C is a light-independent offset and Θ some modeling noise, which is a collection of all other noise sources mostly random in nature (e.g. readout noise, shot noise or photonic noise, quantization noise, etc.). Since all pixels are independent and all operations element-wise, the matrix-elements $y_{x,y} \in Y$ are denoted as $y \in Y$ for simplicity reasons. The same applies to $i \in I$, $k \in K$, $d \in D$, $c \in C$ and $\theta \in \Theta$.

The extraction of the PRNU noise residuals is performed as indicated by Fridrich in [42]. For each image I the noise residual W_I is estimated:

$$W_I = I - F(I) \tag{21.2}$$

where F is a denoising function filtering out the sensor pattern noise. We used two different denoising techniques to extract the PRNU from the images: The wavelet-based denoising filter as described in Appendix A of [43] and the BM3D filter proposed in [44], which is reported to produce better and more consistent results in filtering out the PRNU in [45]. The extracted PRNU noise residual is then normalised in respect to the L_2-norm because its embedding strength is varying between different sensors as explained by [9]. As additional post processing steps a zero mean operation is applied to each extracted PRNU noise residual to suppress artifacts with regular grid structure.

To reduce the PRNU contamination effect from scene details, we apply an image content attenuating PRNU enhancement technique (Model 3 in [39]), subsequently denoted as *ELi*.

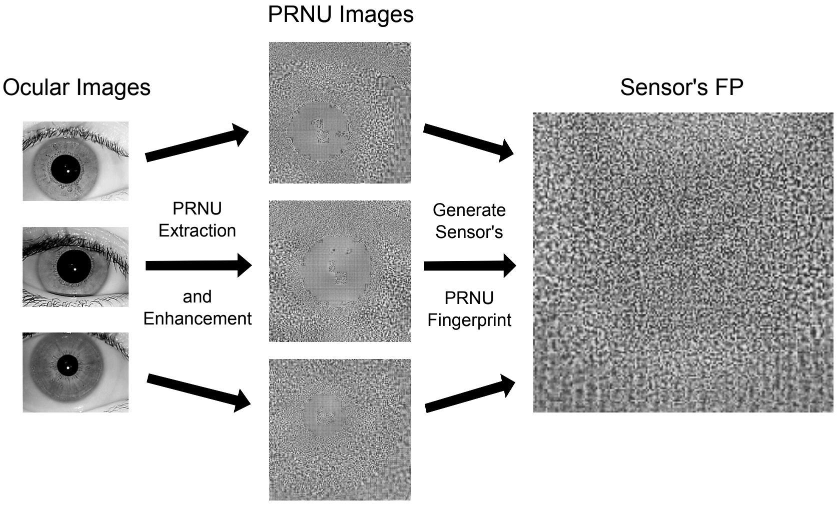

PRNU Images
Ocular Images
Sensor's FP
PRNU Extraction and Enhancement
Generate Sensor's PRNU Fingerprint

Figure 21.2: PRNU noise residual extraction and fingerprint generation with multiple iris images of the same sensor.

Estimating a sensor's PRNU from a single image is usually not sufficient, because that specific image may contain various kinds of disturbances as modeled by Θ in Equation (21.1). Thus multiple images from the same sensor are averaged to isolate the systematic components of all images and suppress these random noise components, as shown in Figure 21.2. This averaged noise is denoted as PRNU fingerprint or reference pattern noise (RPN) in literature. The PRNU fingerprint $\hat{K}$ of a sensor is then estimated using a maximum likelihood estimator for images I_i with $i = 1...N$.

$$\hat{K} = \frac{\sum_{i=1}^{N} W_I^i I^i}{\sum_{i=1}^{N} (I^i)^2} \tag{21.3}$$

The PRNU fingerprint is enhanced using a Wiener filter applied in the DFT domain to suppress periodic artifacts as described in [46].

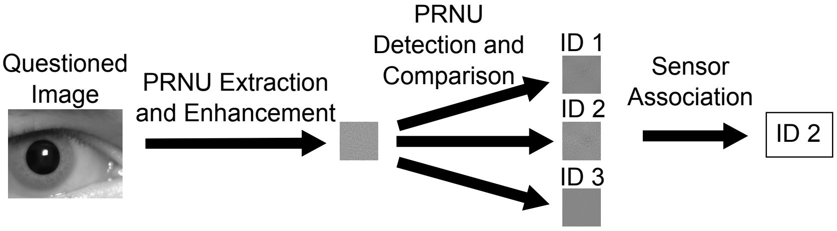

Questioned Image
PRNU Extraction and Enhancement
PRNU Detection and Comparison
ID 1
ID 2
ID 3
Sensor Association
ID 2

Figure 21.3: PRNU noise residual extraction and identification of corresponding sensor.

To determine if an image has been acquired with a specific sensor, the presence of a sensor's PRNU fingerprint in the questioned image has to be detected. Since images acquired with iris sensors are usually not geometrically transformed, this can be done by means of calculating the normalised Cross Correlation (NCC):

$$\rho_{[J,\hat{K}]} = NCC(W_J, J\hat{K}) \tag{21.4}$$

where ρ indicates the correlation between the PRNU residual W_j of the image J and the fingerprint $\hat{K}$ weighted by the image content of J.

An alternative correlation measure to detect the presence of a PRNU fingerprint $\hat{K}$ in an Image I is the Peak Correlation Energy (PCE), proposed by Fridrich in [46]. Fridrich notes that with the PCE the detection threshold will not vary as much as for NCC detector with varying signal length, different cameras and their on-board image processing. It is applied like the NCC detector:

$$\rho_{[J,\hat{K}]} = PCE(W_J, J\hat{K}) \tag{21.5}$$

A schematic illustration for the detection of the correct PRNU fingerprint in a questioned image is given in Figure 21.3.

21.1.2 Iris Texture Classification (ITC)

The input for the Iris Texture Classification (ITC) approach are the preprocessed, segmented, unrolled and normalised iris images originating from various different iris datasets and the output of the classifier is a prediction of the iris sensor used to capture the image or the dataset where the input iris image belongs to, respectively. As the ITC is SVM based, a training phase is needed prior to the use of the classifier, similar to generating a PRNU fingerprint for the PSI approach. In the following the three chosen feature extraction methods, namely DenseSIFT, DMD and LBP are briefly explained. Then the classification approach using a GMM, Fisher Vector encoding and a SVM classifier is described.

21.1.2.1 Feature Extraction

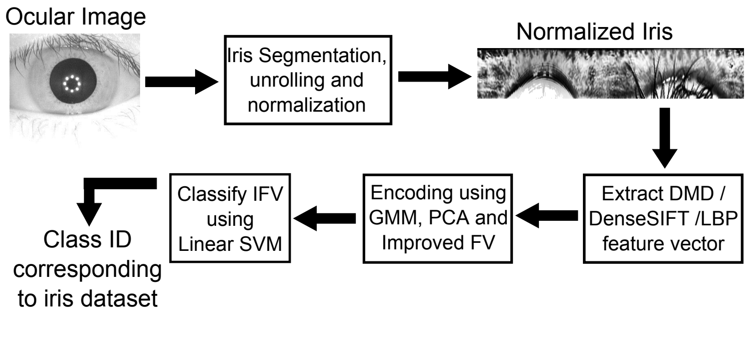

Ocular Image
Iris Segmentation, unrolling and normalization
Normalized Iris
Extract DMD / DenseSIFT /LBP feature vector
Encoding using GMM, PCA and Improved FV
Classify IFV using Linear SVM
Class ID corresponding to iris dataset

Figure 21.4: Flowchart of the Iris Texture Classification (ITC) approach.

DenseSIFT: Is a variant of SIFT. SIFT, the scale invariant feature transform, is a general purpose feature extraction technique used in object recognition proposed

by Lowe [47]. It is invariant to image scale and rotation and robust against various affine distortions, addition of noise, illumination changes and changes of the viewpoint. SIFT locates extrema in the scale-space, localises keypoints, determines their dominant orientation and finally constructs a local descriptor for the keypoint based on a region around it. Fei-Fei et al. [48] proposed to use the local SIFT descriptors on a predefined grid defined across the whole image instead of localising their positions according to scale space extrema. This approach is known as dense SIFT. A 128-dimensional SIFT feature vector is extracted each 3 pixels in 5 different scales (2^0, $2^{-1/2}$, 2^{-1}, $2^{-3/2}$, 2^{-2}). The spatial bins of the SIFT feature descriptor histogram consist of 4 bins in x, 4 bins in y and 8 orientation bins. vl_feat's (`http://www.vlfeat.org`) implementation of DenseSIFT is utilised.
DMD: Dense Micro-block difference is a local feature extraction and texture classification technique proposed by Mehta and Egiazarian [49]. It captures the local structure from image patches (9×9 to 15×15 pixels) at high scales. Instead of the pixels, small blocks of the image which capture the micro-structure are processed. Therefore the pairwise intensity differences of smaller blocks (e.g. 2×2 or 3×3 pixel blocks) calculated in several different directions (not only radial like in LBP) in combination with the average intensity of the whole patch are used to encode the local structure of the patch. Difference values of block pairs located near the centre of the patch are given higher weights than blocks towards the patch boundaries. This should be able to capture the repetitively characteristic local structure providing discriminative information.
LBP: The local binary patterns proposed by Ojala [50] observe the variations of pixels in a local neighborhood. These variations are thresholded by the central pixel value to obtain a binary decision, which is then encoded as a scalar value. The occurrences of each scalar value for all pixels in the image are represented in a histogram, which forms the extracted feature vector.

21.1.2.2 Feature Encoding

We utilize the Improved Fisher Vector Encoding (IFV) scheme [51] in the same way as it is done in [52, 49]. IFV is usually used in object recognition. Fisher vector encoding starts by extracting local SIFT descriptors densly (DenseSIFT) and at multiple scales to get a feature vector f. We not only use DenseSIFT features but also DMD and LBP ones as input for the next steps. The feature vector f is then soft-quantised using a Gaussian Mixture Model (GMM) with K modes where the Gaussian covariance matrices are assumed to be diagonal. The local descriptors present in f are first decorrelated and then dimensionality reduced (optional) by PCA. So far this describes the standard Fisher Vector encoding [53]. The IFV now adds signed square rooting and l^2 normalization as described in [51].

21.1.2.3 Classification

A linear SVM is then used to classify the IFV encoded features. We experimented with different types of kernels $K(x',x'')$ (linear, Hellinger, exponential) and the linear kernel lead to the most promising results. The input data to the SVM (IFV encoded feature vectors) is normalised such that $K(x',x'') = 1$ which usually improves the performance. The SVM is trained using a standard non-linear SVM solver on a subset of the unrolled, normalised iris images which is subsequently not used for the testing (evaluation) step.

21.2 Datasets

To enable a meaningful comparision with the previous work of El-Naggar and Ross [41] we attempted to use the same iris datasets they originally used and extend the number of datasets. Unfortunately we were not able to acquire the MGBC and the WVU iris dataset. Thus we use the remaining 6 datasets they used plus 3 additional iris dataset which are described in the following. All of them are publicly available, common datasets which have been utilised in many different iris recognition related works. Figure 21.5 shows some example images for each of the datasets. Table 21.1 summarizes the most important attributes of the datasets. In the following a short description of each single dataset is given:

CASIA V2: We use the first subset of the CASIA V2 iris database [54] (device 1). This subset consists of 1200 images and was captured using an OKI Irispass-h sensor by the Chinese Academy of Sciences Institue of Automation (CASIA).
CASIA V3: was again captured by the Chinese Academy of Sciences Institue of Automation (CASIA) [54] and consists of several different subsets. We used the CASIA V3 Interval subset in accordance with the work of El-Naggar and Ross. This subset consists of 2639 images captured with a self-developed close-up iris camera.
CASIA V4: This is the V4 version of the iris dataset provided by CASIA [54]. Again it consists of several subsets, where we used the Thousands subset. It consists of 20000 images which were collected using an IrisKing IKEMB-100 camera.
ICE2005: NIST, the National Institute of Standards and Technology in the US conducted a series of biometric recognition contests, one of them was the Iris Challenge Evaluation (ICE) in 2005. The ICE2005 [55] images where captured at the University of Notre Dame with a LG EOU 2200 iris camera and consists of 2953 images.
IITD: The IIT Delhi Iris Database [56] consists of 1120 images and was acquired by the Biometrics Research Laboratory in the Indian Institute of Technology Delhi (IITD) in 2007. The images were captures with an JIRIS JPC1000 digital CMOS iris camera.
MMU2: The MMU V2 iris database [57] consists of 995 iris images. These images are collected using a Panasonic BM-ET100US Authenticam.
UBIRIS: The Noisy Visible Wavelength Iris Image Database UBIRIS V1 [58] consists of 1877 images collected in 2004. The images were captured with a Nikon E5700 digital camera in two sessions.
UPOL: The Univerzita Palackho v Olomouci iris dataset [59] consists of 384 images. The irises were scanned by TOPCON TRC50IA optical device connected with SONY DXC-950P 3CCD camera.
UTIRIS: University of Tehran IRIS (UTIRIS) image dataset [60] consists of two different sessions, one captured using visible wavelength illumination and the other one using near-infrared illumination. We only used the near infrared subset, which consists of 793 images. The infrared images were captured with an ISG Lightwise LW iris camera.

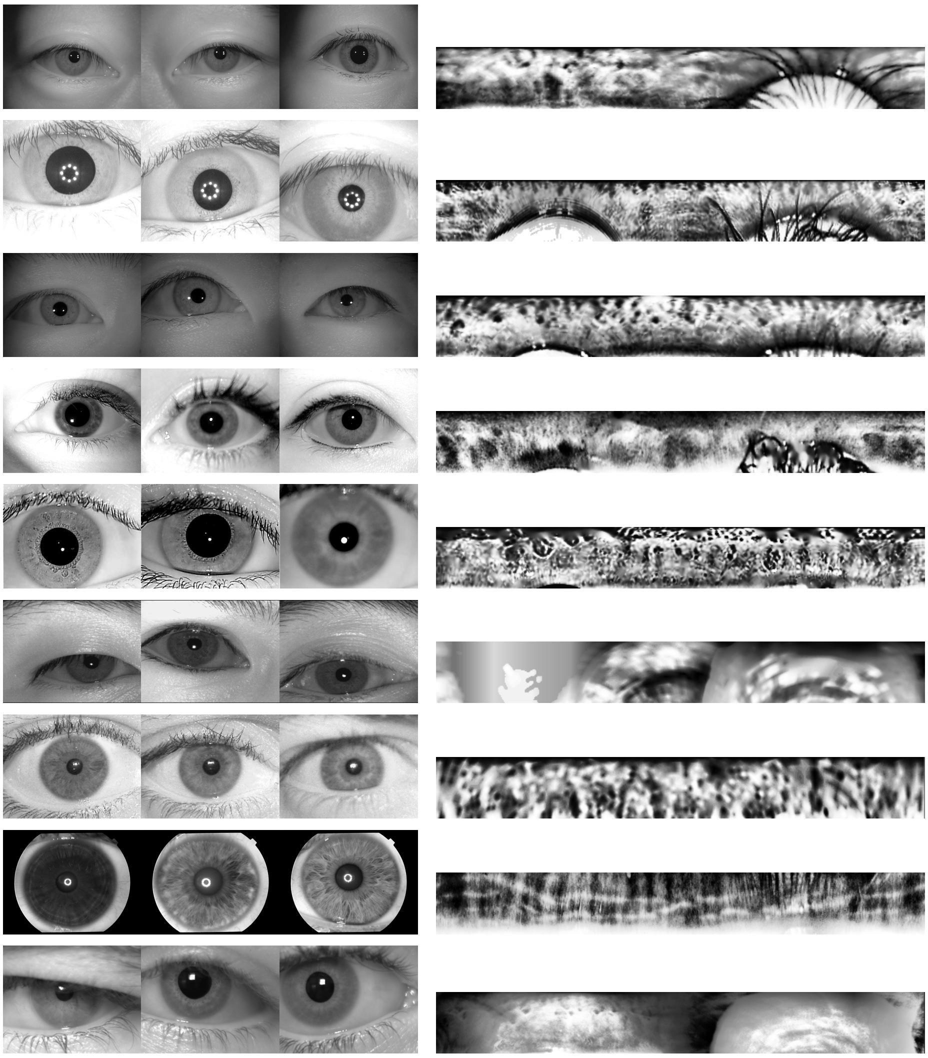


Figure 21.5: Ocular image and normalised iris image samples from different datasets, from top to bottom: CASIA V2, CASIA V3, CASIA V4, ICE2005, IITD, MMU2, UBIRIS, UPOL, UTIRIS

Dataset	# IMG	Sensor	Illumination	Resolution	Class ID
CASIA V2	1200	OKI IRISPASS-h	near infrared	480 × 640	1
CASIA V3	2639	CASIA Iris camera	near infrared	320 × 280	2
CASIA V4	20000	IrisKing IKEMB-100	near infrared	640 × 480	3
ICE2005	2953	LG EOU 2200 iris camera	near infrared	480 × 640	4
IITD	1120	JIRIS, JPC1000 camera	near infrared	240 × 320	5
MMU2	995	Panasonic BM-ET100US Authenticam	near infrared	320 × 238	6
UBIRIS	1877	Nikon E5700	natural lighting	200 × 150	7
UPOL	384	SONY DXC-950P 3CCD camera	camera flash	768 × 576	8
UTIRIS	793	ISG Lightwise LW	near infrared	1000 × 776	9

Table 21.1: Attributes of iris datasets.

21.3 Experimental Setup

We follow the same test methodology as El-Naggar and Ross [41]. For the Iris Texture Classification (ITC) approach as described in Section 21.2 each dataset is randomly split into two distinct subsets, a training and a testing one. UPOL is the iris dataset containing the least images, 384 images only, thus it is split 50:50 into 192 training and 192 testing images. Consequently, for all other images also 192 training and 192 testing images are chosen for the corresponding subsets. The first step in the processing chain is the preprocessing of the ocular images, including iris segmentation and iris unrolling. The unrolled iris patches are then normalised and all having a size 512 × 64 pixels. This is done utilizing the USIT (University of Salzburg Iris Toolkit, Version 2.0 available at `http://www.wavelab.at/sources/USIT/`) software toolkit in version 1.0.3. For the segmentation step the WAHET (Weighted Adaptive Hough and Ellipsopolar Transform) method is used. Figure 21.5 shows one example of an unrolled and normalised iris image for each dataset. For further details on the exact implementation of WAHET and the iris unrolling the interested reader is referred to [61]. The next step is the feature extraction using DenseSIFT, DMD and LBP. Afterwards, the features are dimensionality reduced using a GMM and then Fisher Vector encoding is applied before they are put into a linear SVM for classification. A 5-fold cross validation is performed and the mean results of all 5 runs are used as final results shown below.

For the PRNU based Sensor Identification (PSI) approach we decided to extract the PRNU from a central patch with varying sizes ranging from 64 × 64 up to 576 × 576 pixels because of the varying image size of the data sets. We furthermore evaluate the effect of applying the content attenuation PRNU enhancement *ELi*, described in section 21.1.1, in contrast to not applying it. The configurations for the extraction and post-processing of the PRNU are: Extracted PRNU sizes (from 64 × 64 up to 576 × 576 pixels), denoising filters (Wavelet and BM3D), PRNU enhancements (ELi and NoEnh) and PRNU detectors (NCC and PCE). Due to the different image sizes of the datasets the number of sensors to discriminate decreases for an increasing PRNU size as shown by the value in parentheses next to the PNRU size in Table 21.8. For each run we selected 192 random images for each data set for the

Method	DenseSIFT	DMD	LBP
mAcc	0.9838	0.9688	0.8715
mAP	0.9968	0.9878	0.9172

Table 21.2: Mean accuracies (mACC) and mean average precisions (mAP) for DenseSIFT, DMD and LBP

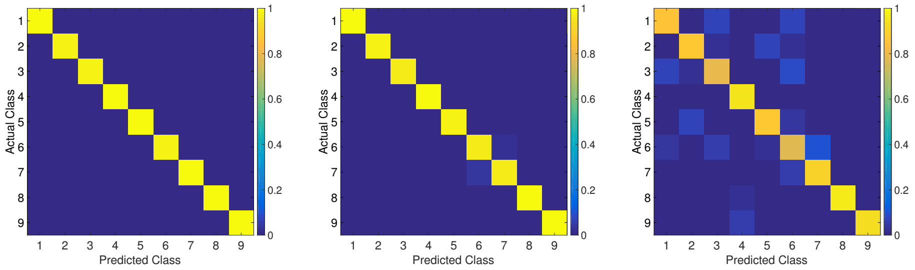


Figure 21.6: Confusion matrix for DenseSIFT, DMD and LBP

generation of the PRNU fingerprint ("training" set) and another 192 random images as the test set, without overlapping images between both sets. We compute the NCC and PCE correlation scores for all test images with all generated PRNU fingerprints, where the predicted sensor (or class) is determined by means of the highest (rank one) correlation score. The larger the size of the extracted PRNU, the less sensors could be used for evaluating the identification performance for the sensors. The experiment was repeated 5 times (5-fold cross validation), where the final result is the average of all 5 runs. The described parameters have been chosen to make the results of both identification/classification approaches as comparable as possible.

21.4 Experimental Results

Table 21.2 summarises the results of the ICT approach. It lists the mean accuracy (mAcc) as well as the mean average precision (mAP) over all 5 runs. The accuracy describes the number of correctly classified items (true positives + true negatives) over the number of total items per class calculated per class. The mAcc is just the mean over all single accuracies. The average precision (AP) describes the area under the precision/recall curve calculated per query/class. The mAP is the mean over all AP values. It can be clearly seen that the ICT approach works best using DenseSIFT features. Using DMD and LBP the recognition performance both in terms of the mACC and the mAP is still clearly over 90%. The results show that the ICT approach is able to determine the source of an unrolled iris texture image with a very high accuracy considering the nine iris datasets.

Figure 21.6 shows the confusion matrices for ICT. The numbers on the axes are corresponding to the class IDs in table 21.1. Considering DenseSIFT it can be seen

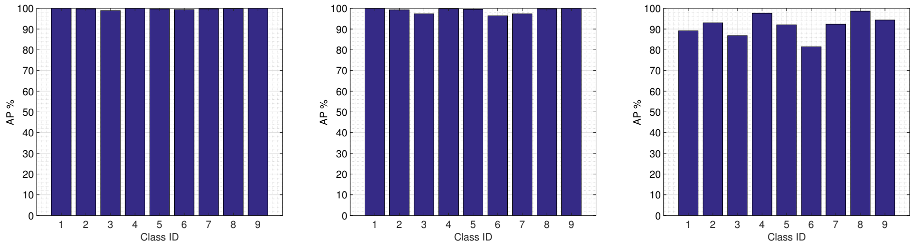


Figure 21.7: Average precision for DenseSIFT, DMD and LBP

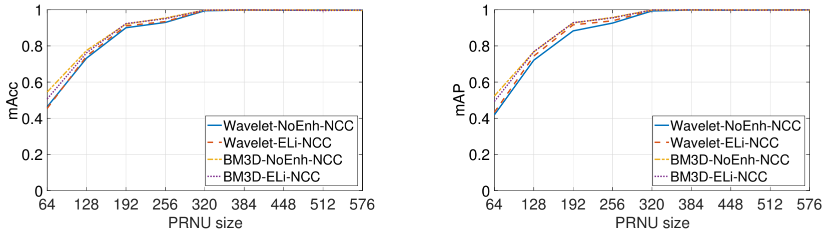


Figure 21.8: Mean accuracy (mAcc) and mean average precision (mAP) for selected PRNU patch sizes and PRNU extraction configurations.

that for CASIA V2, ICE2005, MMU2, UBIRIS, UPOL and UTIRIS all images are correctly classified as belonging to the actual dataset. Only some of the CASIA V3, CASIA V4 and IITD images are misclassified.

Figure 21.7 shows the average precision plots for ICT. Again it can be seen that classification works perfectly for CASIA V2, MMU2, UBIRIS, UPOL, ICE2005 and UTIRIS considering DenseSIFT. Considering DMD it still works perfectly for CASIA V2, ICE2005, UPOL and UTIRIS but no longer for MMU2 and UBIRIS though still quite acceptably. LBP's performance is a bit worse.

Table 21.3 lists the PSI results which show that the PRNU size affects the identification performance most across all configurations. Reasonable mAcc and mAP rates can already be achieved with 192 × 192 pixel patches, while a patch size larger than 320 × 320 yields very good results for the identification of the different iris sensors through their PRNU fingerprint. Neither the choice of PRNU detector nor PRNU enhancement makes a big difference in this case, but better results can be achieved by choosing the BM3D denoising filter over the Wavelet filter for smaller PRNU sizes, as shown in Figure 21.8.

Next we are having a closer look at the results for the single classes or sensors. Figure 21.9 shows the confusion matrix for both denoising filters using a small patch size of 64 × 64 pixels, no content attenuation PRNU enhancement (NoEnh) and the NCC detector. It can be seen that the identification performance varies highly among the different classes, where class 9 shows very good results and the classes 1, 2 and 7 show very low identification performance independent of the denoising filter. All other classes show higher accuracies when the BM3D filter is used.

		Wavelet				BM3D			
	PRNU	NoEnh		ELi		NoEnh		ELi	
	Size	NCC	PCE	NCC	PCE	NCC	PCE	NCC	PCE
mAcc	64 (9)	0.4633	0.4587	0.4545	0.4456	0.5451	0.5397	0.5068	0.4845
	128 (9)	0.7326	0.7300	0.7437	0.7380	0.7752	0.7696	0.7620	0.7554
	192 (8)	0.9007	0.8932	0.9112	0.9008	0.9210	0.9279	0.9237	0.9197
	256 (6)	0.9300	0.9358	0.9347	0.9326	0.9545	0.9507	0.9505	0.9446
	320 (5)	0.9946	0.9963	0.9973	0.9988	0.9981	0.9994	0.9981	0.9983
	384 (5)	0.9988	0.9987	0.9990	0.9990	0.9983	0.9981	0.9992	0.9992
	448 (5)	0.9973	0.9988	0.9983	0.9992	0.9990	0.9998	0.9983	0.9971
	512 (2)	0.9990	0.9974	0.9984	0.9932	0.9984	0.9990	0.9958	0.9943
	576 (2)	0.9984	0.9979	0.9974	0.9995	0.9964	0.9974	0.9984	0.9964
mAP	64 (9)	0.4184	0.4033	0.4325	0.4135	0.5227	0.5146	0.4925	0.4691
	128 (9)	0.7209	0.7114	0.7438	0.7352	0.7674	0.7619	0.7675	0.7572
	192 (8)	0.8832	0.8844	0.9166	0.9092	0.9271	0.9300	0.9287	0.9236
	256 (6)	0.9264	0.9268	0.9389	0.9384	0.9576	0.9530	0.9549	0.9506
	320 (5)	0.9934	0.9949	0.9977	0.9986	0.9985	0.9992	0.9989	0.9986
	384 (5)	0.9989	0.9989	0.9992	0.9990	0.9988	0.9987	0.9994	0.9994
	448 (5)	0.9976	0.9988	0.9985	0.9988	0.9991	0.9994	0.9986	0.9977
	512 (2)	0.9991	0.9984	0.9987	0.9966	0.9990	0.9997	0.9977	0.9973
	576 (2)	0.9995	0.9989	0.9990	0.9998	0.9982	0.9991	0.9994	0.9982

Table 21.3: Mean accuracy (mAcc) and mean average precision (mAP) for all tested PRNU patch sizes and PRNU extraction configurations. The number in parentheses next to the PRNU size indicates the number of different sensors.

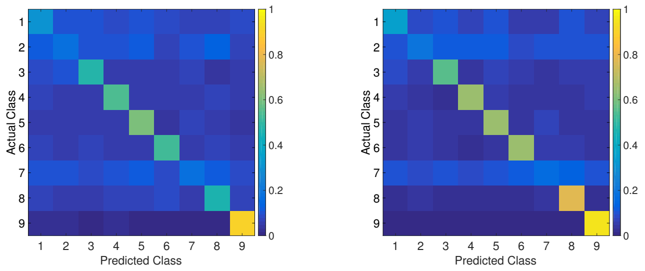


Figure 21.9: Confusion matrices using the Wavelet (left) and BM3D (right) denoising filters for 64×64 pixels PRNU patch size, no content attenuation PRNU enhancement (NoEnh) and NCC detector.

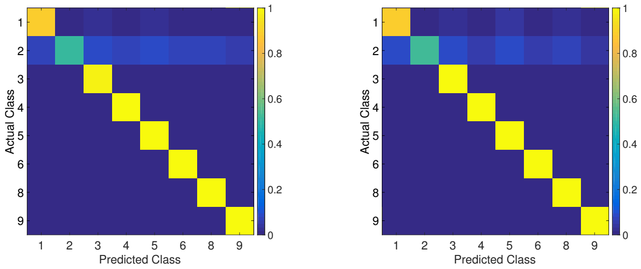


Figure 21.10: Confusion matrices using the NCC (left) and PCE (right) detectors for 192 × 192 *pixels PRNU patch size, ELi content attenuation PRNU enhancement and BM3D denoising filter.*

TS size	DenseSIFT	DMD	LBP	PRNU128	PRNU256	PRNU512
192	0.9937	0.9810	0.9116	0.8227	0.9540	0.9999
96	0.9919	0.9547	0.8143	0.7870	0.9348	0.9979
48	0.9833	0.9564	0.5038	0.7426	0.9069	0.9978
24	0.9668	0.9277	-	0.7073	0.8594	0.9937
12	0.9367	0.8805	-	0.6363	0.8003	0.9796
6	0.8766	0.8062	-	0.5244	0.7476	0.9565
3	0.7897	0.6921	-	0.4817	0.6905	0.9348
1	0.6320	0.5749	-	0.3297	0.5734	0.8623

Table 21.4: Mean average precisions (mAP) for DenseSIFT, DMD, LBP, and PRNU with sizes 128 × 128, 256 × 256 *and* 512 × 512 *for different training set sizes (TS size).*

Having a look at a larger PNRU size of 192 × 192 pixels, BM3D denoising filter and ELi PRNU enhancement, as shown in Figure 21.10, reveals that the identification performance for both detectors, NCC and PCE, is practically identical with only slight differences for all classes. This indicates that the choice of detector is not critical for the overall performance.

The ITC approach in general and the PSI approach with PRNU sizes at least 512 × 512 pixels outperform the approach by El-Naggar and Ross [41].

21.5 Practical Discussion

In order to secure a biometric system against insertion attacks (described in the introduction) the authenticity of the biometric samples, i.e. images, has to be verified. We examined two approaches tailored to identify the sensor an iris image was captured with. Both approaches can be used for existing biometric systems and while setting-up new ones because they rely solely on intrinsic image properties. The first one is based on the PRNU of the iris sensors and the second one on texture features. We examined different training set sizes. The results shown in Table 21.4, indicate that both achieve good classification results if some conditions are met. The ITC ap-

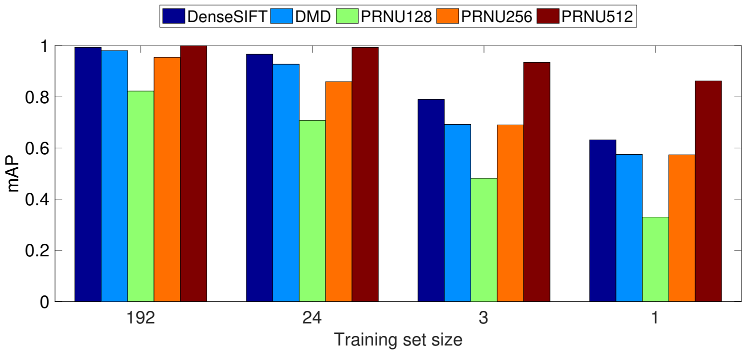


Figure 21.11: Exemplary mean average precision (mAP) scores for selected training set sizes.

proach works well for a broad range of image resolutions as long as there are enough training images available, i.e. at least 10 training images should be available. Using the LBP feature extractor for the ITC approach causes the training to fail for smaller training set sizes, as shown by missing values in Table 21.4. In these cases the LBP feature vectors are not distinctive enough and cannot be soft-quantized by the GMM. The different image resolutions and iris sizes in the images among the datasets require varying unrolling parameters. Therefore, unrolling and normalisation cause a separate level of interpolation for each dataset. Due to the texture classification nature of the ITC approach the results could be biased by the interpolation artifacts that may improve the discriminative power of the datasets but not the sensors themselves. This effect could eventually be mitigated by using the ocular images as input in combination with different features like BSIF [62], which is designed to capture image characteristics similar to the PRNU.

The PSI approach works well for bigger PRNU sizes and also works for very few training images, c.f. it even works with one single training image for the PRNU512 case. The PRNU is extracted directly from the ocular images as opposed to the unrolled iris texture in the ITC approach. Thus no additional bias is introduced and the discrimination relies solely on the sensors' characteristics.

In this work we only investigated different types of sensors, but not multiple sensors of the same model and manufacturer. As mentioned in the introduction the PRNU is able to distinguish between specific sensor instances of the same model, which is an advantage over the ITC approach in practical deployments since the attacker may have access to the same sensor model. It needs to be clarified whether the ITC approach is able to handle this kind of set-up as well.

Both approaches are suitable if it comes to securing a biometric system depending on the system's configuration. For biometric systems dealing with smaller images but with many training images available the ITC approach is favourable, for systems dealing with larger images but only very few training images available the PSI approach should be used. This is further illustrated in Figure 21.11. If only few training images are available and the images are small then a fusion of the two approaches could improve the results. The image size cannot be changed easily but it is easy to provide some more training images (just capture additional data with the sensor) and thus the ITC approach can be used again. For securing a biometric system at first the respective approach has to be trained (PRNU fingerprint generation for the PSI approach) using images of the biometric sensor(s). Every time a new biometric sample is captured the image is analysed using the pretrained classifier which then tells if the image was captured by one of the biometric sensors it was trained with or not. In the latter case it is very likely that an insertion attack happened and the authentication process is aborted.

21.6 Conclusion

In this chapter we examined two passive approaches to secure an iris recognition system against insertion attacks by verifying the authenticity of the iris images. The first one, named PSI, is based on the photo response non-uniformity (PRNU) of image sensors and the second one, named ITC, exploits the texture information of unrolled iris images. The examination was performed using images from 9 distinct iris databases or sensors, respectively.

The results show that both approaches perform well in identifying the correct sensor an iris image was captured with, though the performance of the PSI approach is dependent on the size of the extracted PRNU. The ITC approach worked well for all datasets. We furthermore examined the impact of the number of images available for the training of both approaches.

Each approach has its advantages and drawbacks depending on the configuration of the biometric system: The PSI approach gives better results if only a small number of high-resolution images is available, while the ITC approach needs a higher number of images to achieve an acceptable performance, but the advantage is they do not need to be of high resolution. In addition the PSI approach is also suited to distinguish different sensors of the same model. This helps in detecting an injection of images from the same sensor model as deployed in the biometric system. If only a small number of low resolution images is available, a fusion of both approaches is likely to improve the overall performance.

Bibliography

[1] J. Daugman, “How iris recognition works,” *International Conference on Image Processing*, vol. 1, pp. I–33–I–36, 2002.

[2] N. Ratha, J. Connell, and R. Bolle, “Enhancing security and privacy in biometrics-based authentication systems,” *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[3] H. T. Sencar, M. Ramkumar, and A. N. Akansu, *Data hiding fundamentals and applications. Content security in digital multimedia.*, 2004.

[4] C. Wu and C. Kuo, “Comparison of two speech content authentication approaches,” in *Proceedings of SPIE, Security and Watermarking of Multimedia Contents IV*, vol. 4675, San Jose, CA, USA, 2002.

[5] M. Schneider and S.-F. Chang, “A robust content based digital signature for image authentication,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP’96)*, Lausanne, Switzerland, 1996.

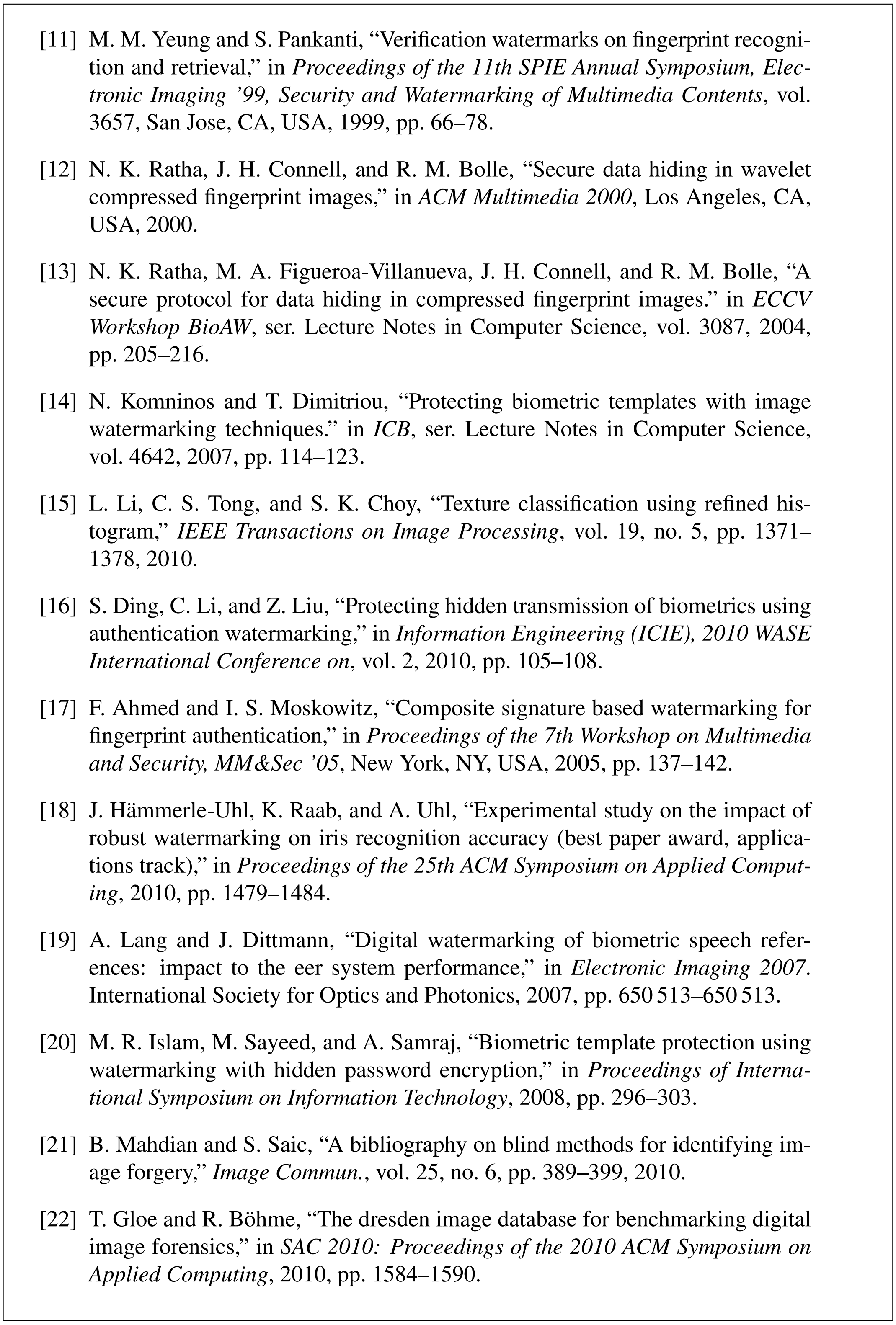
[6] J. Tzeng, W.-L. Hwang, and I. Chern, “Enhancing image watermarking methods by second order statistics,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP’01)*, Thessaloniki, Greece, 2001.

[7] C.-S. Lu and H.-Y. M. Liao, “Oblivious watermarking using generalized gaussian,” in *Proceedings of the 7th International Conference on Fuzzy Theory and Technology*, Atlantic City, NJ, USA, 2000, pp. 260–263.

[8] W.-K. Lin and N. Burgess, “Listless zerotree coding for color images,” in *32nd Asilomar Conference on Signals, System and Computers*, CA, USA, 1998.

[9] A. Uhl and Y. Höller, “Iris-sensor authentication using camera PRNU fingerprints,” in *Proceedings of the 5th IAPR/IEEE International Conference on Biometrics (ICB’12)*, New Delhi, India, 2012, pp. 1–8.

[10] J. Hämmerle-Uhl, K. Raab, and A. Uhl, “Watermarking as a means to enhance biometric systems: A critical survey,” in *Proceedings of the 2011 Information Hiding Conference (IH’11)*, ser. Springer LNCS, vol. 6958, Prague, Czech Republic, 2011, pp. 238–254.



[11] M. M. Yeung and S. Pankanti, “Verification watermarks on fingerprint recognition and retrieval,” in *Proceedings of the 11th SPIE Annual Symposium, Electronic Imaging ’99, Security and Watermarking of Multimedia Contents*, vol. 3657, San Jose, CA, USA, 1999, pp. 66–78.

[12] N. K. Ratha, J. H. Connell, and R. M. Bolle, “Secure data hiding in wavelet compressed fingerprint images,” in *ACM Multimedia 2000*, Los Angeles, CA, USA, 2000.

[13] N. K. Ratha, M. A. Figueroa-Villanueva, J. H. Connell, and R. M. Bolle, “A secure protocol for data hiding in compressed fingerprint images.” in *ECCV Workshop BioAW*, ser. Lecture Notes in Computer Science, vol. 3087, 2004, pp. 205–216.

[14] N. Komninos and T. Dimitriou, “Protecting biometric templates with image watermarking techniques.” in *ICB*, ser. Lecture Notes in Computer Science, vol. 4642, 2007, pp. 114–123.

[15] L. Li, C. S. Tong, and S. K. Choy, “Texture classification using refined histogram,” *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1371–1378, 2010.

[16] S. Ding, C. Li, and Z. Liu, “Protecting hidden transmission of biometrics using authentication watermarking,” in *Information Engineering (ICIE), 2010 WASE International Conference on*, vol. 2, 2010, pp. 105–108.

[17] F. Ahmed and I. S. Moskowitz, “Composite signature based watermarking for fingerprint authentication,” in *Proceedings of the 7th Workshop on Multimedia and Security, MM&Sec ’05*, New York, NY, USA, 2005, pp. 137–142.

[18] J. Hämmerle-Uhl, K. Raab, and A. Uhl, “Experimental study on the impact of robust watermarking on iris recognition accuracy (best paper award, applications track),” in *Proceedings of the 25th ACM Symposium on Applied Computing*, 2010, pp. 1479–1484.

[19] A. Lang and J. Dittmann, “Digital watermarking of biometric speech references: impact to the eer system performance,” in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 650 513–650 513.

[20] M. R. Islam, M. Sayeed, and A. Samraj, “Biometric template protection using watermarking with hidden password encryption,” in *Proceedings of International Symposium on Information Technology*, 2008, pp. 296–303.

[21] B. Mahdian and S. Saic, “A bibliography on blind methods for identifying image forgery,” *Image Commun.*, vol. 25, no. 6, pp. 389–399, 2010.

[22] T. Gloe and R. Böhme, “The dresden image database for benchmarking digital image forensics,” in *SAC 2010: Proceedings of the 2010 ACM Symposium on Applied Computing*, 2010, pp. 1584–1590.

[23] M. K. Johnson and H. Farid, “Exposing digital forgeries in complex lighting environments,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 450–461, 2007.

[24] K. Choi, E. Lam, and K. Wong, “Automatic source camera identification using the intrinsic lens radial distortion,” *OPTICS EXPRESS*, vol. 14, no. 24, pp. 11 551–65, 2006.

[25] Z. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh, “Methods for identification of images acquired with digital cameras,” in *Proceedings of SPIE, Enabling Technologies for Law Enforcement and Security*, vol. 4232, 2001, pp. 505–512.

[26] M. Kharrazi, H. T. Sencar, and N. Memon, “Blind source camera identification,” in *Image Processing, 2004. ICIP ’04. 2004 International Conference on*, vol. 1, 2004, pp. 709–712 Vol. 1.

[27] S. Bayram, H. Sencar, N. Memon, and I. Avcibas, “Source camera identification based on CFA interpolation,” in *Proceedings of the IEEE International Conference on Image Processing, ICIP ’05*, vol. 2, Genoa, Italy, 2005, pp. 69–72.

[28] A. De Rosa, A. Piva, M. Fontani, and M. Iuliani, “Investigating multimedia contents,” in *Security Technology (ICCST), 2014 International Carnahan Conference on*, 2014, pp. 1–6.

[29] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*, 2009.

[30] N. Bartlow, N. Kalka, B. Cukic, and A. Ross, “Identifying sensors from fingerprint images,” in *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, 2009, pp. 78–84.

[31] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, “Determining image origin and integrity using sensor noise,” *IEEE Transactions on Information Security and Forensics*, vol. 3, no. 1, pp. 74–90, 2008.

[32] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, “A bayesian-mrf approach for prnu-based image forgery detection,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 554–567, 2014.

[33] M. Goljan, J. Fridrich, and T. Filler, “Large scale test of sensor fingerprint camera identification,” in *Proceedings of SPIE, Electronic Imaging, Security and Forensics of Multimedia Contents XI*, San Jose, CA, USA, 2009.

[34] M. Goljan, J. Fridrich, and T. Filler, “Managing a large database of camera fingerprints,” in *Proceedings of SPIE, Media Forensics and Security XII*, San Jose, CA, USA, 2010.

[35] K. Rosenfeld and H. Sencar, “A study of the robustness of prnu-based camera identification,” in *Proceedings of SPIE, Media Forensics and Security XI*, vol. 7254, San Jose, CA, USA, 2009, pp. 72 540M – 725 408M.

[36] E. Alles, Z. Geradts, and C. Veenman, “Source camera identification for heavily jpeg compressed low resolution still images,” *Journal of Forensic Sciences*, vol. 54, no. 3, pp. 628–638, 2009.

[37] M. Goljan, J. Fridrich, and M. Chen, “Defending against fingerprint-copy attack in sensor-based camera identification,” *IEEE Transactions on Information Security and Forensics*, vol. 6, no. 1, pp. 227–236, 2011.

[38] N. Kalka, N. Bartlow, B. Cukic, and A. Ross, “A preliminary study on identifying sensors from iris images,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2015.

[39] C.-T. Li, “Source camera identification using enhanced sensor pattern noise,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, 2010.

[40] M. Goljan, J. Fridrich, and M. Chen, “Sensor noise camera identification: Countering counter-forensics,” in *Proceedings of SPIE, Media Forensics and Security XII*, San Jose, CA, USA, 2010.

[41] S. El-Naggar and A. Ross, “Which dataset is this iris image from?” in *Information Forensics and Security (WIFS), 2015 IEEE International Workshop on*. IEEE, 2015, pp. 1–6.

[42] J. Fridrich, “Digital image forensic using sensor noise,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, 2009.

[43] J. Lukas, J. Fridrich, and M. Goljan, “Digital camera identification from sensor pattern noise.” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[44] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising with block-matching and 3d filtering,” in *Electronic Imaging 2006*. International Society for Optics and Photonics, 2006, pp. 606 414–606 414.

[45] A. Cortiana, V. Conotter, G. Boato, and F. D. Natale, “Performance comparison of denoising filters for source camera identification,” in *Media Watermarking, Security, and Forensics XIII*, ser. Proceedings of SPIE, vol. 7880, 2011, p. 788007.

[46] J. Fridrich, “Sensor defects in digital image forensics,” in *Digital Image Forensics: There is more to a picture than meets the eye*, 2012, ch. 6, pp. 179–218.

[47] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[48] L. Fei-Fei and P. Perona, “A bayesian hierarchical model for learning natural scene categories,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 524–531 vol. 2.

[49] R. Mehta and K. Egiazarian, *Texture Classification Using Dense Micro-block Difference (DMD)*, ser. Lecture Notes in Computer Science. Springer-Verlag, Berlin, 2015, pp. 643–658.

[50] T. Ojala, M. Pietikainen, and D. Harwood, “Performance evaluation of texture measures with classification based on kullback discrimination of distributions,” in *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, vol. 1, 1994, pp. 582 –585 vol.1.

[51] F. Perronnin, J. Sánchez, and T. Mensink, “Improving the fisher kernel for large-scale image classification,” in *European conference on computer vision (ECCV10)*. Springer, 2010, pp. 143–156.

[52] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, “Describing textures in the wild,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’14)*, 2014, pp. 3606–3613.

[53] F. Perronnin and C. Dance, “Fisher kernels on visual vocabularies for image categorization,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[54] N. L. of Pattern Recognition. Casia iris v4 database. http://biometrics.idealtest.org/.

[55] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, and W. T. Scruggs, “The iris challenge evaluation 2005,” in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*. IEEE, 2008, pp. 1–8.

[56] A. Kumar and A. Passi, “Comparison and combination of iris matchers for reliable personal authentication,” *Pattern recognition*, vol. 43, no. 3, pp. 1016–1026, 2010.

[57] Cteo, “Mmu2 iris image database,” *Available at: http://pesona.mmu.edu.my/ ccteo/*, 2008.

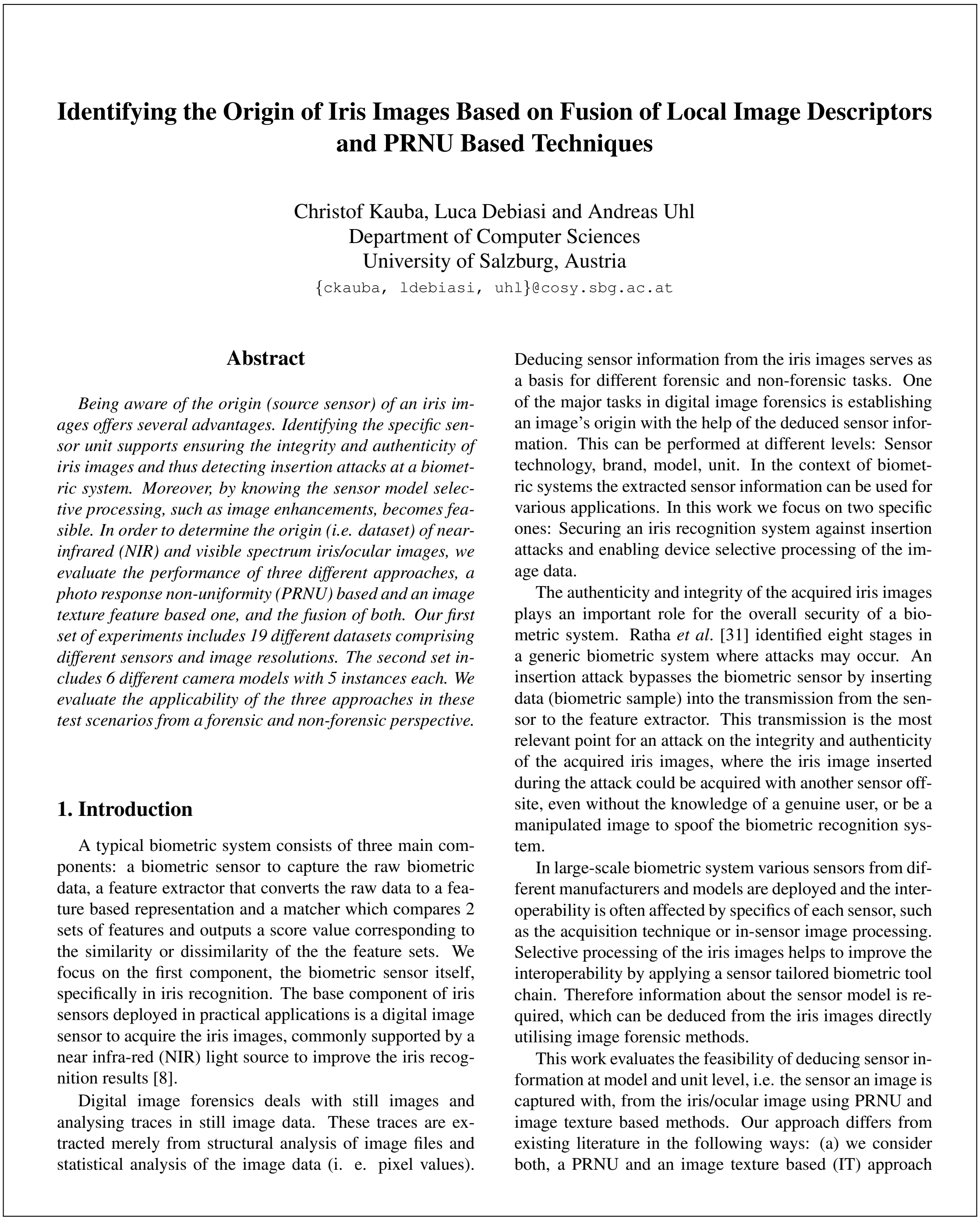
[58] H. Proenca and L. Alexandre, “UBIRIS: a noisy iris image database,” in *Image Analysis and Processing - ICIAP 2005*, ser. Lecture Notes on Computer Science, vol. 3617. Cagliari, Italy: Springer-Verlag, 2005, pp. 970–977.

[59] M. Dobes and L. Machala, “Upol iris image database, 2004,” *Available at: http;//www.phoenix.inf.upol.cz/iris*, 2013.

[60] M. Hosseini, B. Araabi, and H. Soltanian-Zadeh, “Pigment melanin: Pattern for iris recognition,” *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 4, pp. 792 –804, 2010.

[61] C. Rathgeb, A. Uhl, and P. Wild, *Iris Recognition: From Segmentation to Template Security*, ser. Advances in Information Security, 2013, vol. 59.

[62] J. Kannala and E. Rahtu, “BSIF: binarized statistical image features,” in *Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012, Tsukuba, Japan, November 11-15, 2012*, 2012, pp. 1363–1366.



Identifying the Origin of Iris Images Based on Fusion of Local Image Descriptors and PRNU Based Techniques

Christof Kauba, Luca Debiasi and Andreas Uhl
Department of Computer Sciences
University of Salzburg, Austria
{ckauba, ldebiasi, uhl}@cosy.sbg.ac.at

Abstract

Being aware of the origin (source sensor) of an iris images offers several advantages. Identifying the specific sensor unit supports ensuring the integrity and authenticity of iris images and thus detecting insertion attacks at a biometric system. Moreover, by knowing the sensor model selective processing, such as image enhancements, becomes feasible. In order to determine the origin (i.e. dataset) of near-infrared (NIR) and visible spectrum iris/ocular images, we evaluate the performance of three different approaches, a photo response non-uniformity (PRNU) based and an image texture feature based one, and the fusion of both. Our first set of experiments includes 19 different datasets comprising different sensors and image resolutions. The second set includes 6 different camera models with 5 instances each. We evaluate the applicability of the three approaches in these test scenarios from a forensic and non-forensic perspective.

1. Introduction

A typical biometric system consists of three main components: a biometric sensor to capture the raw biometric data, a feature extractor that converts the raw data to a feature based representation and a matcher which compares 2 sets of features and outputs a score value corresponding to the similarity or dissimilarity of the the feature sets. We focus on the first component, the biometric sensor itself, specifically in iris recognition. The base component of iris sensors deployed in practical applications is a digital image sensor to acquire the iris images, commonly supported by a near infra-red (NIR) light source to improve the iris recognition results [8].

Digital image forensics deals with still images and analysing traces in still image data. These traces are extracted merely from structural analysis of image files and statistical analysis of the image data (i. e. pixel values). Deducing sensor information from the iris images serves as a basis for different forensic and non-forensic tasks. One of the major tasks in digital image forensics is establishing an image's origin with the help of the deduced sensor information. This can be performed at different levels: Sensor technology, brand, model, unit. In the context of biometric systems the extracted sensor information can be used for various applications. In this work we focus on two specific ones: Securing an iris recognition system against insertion attacks and enabling device selective processing of the image data.

The authenticity and integrity of the acquired iris images plays an important role for the overall security of a biometric system. Ratha *et al.* [31] identified eight stages in a generic biometric system where attacks may occur. An insertion attack bypasses the biometric sensor by inserting data (biometric sample) into the transmission from the sensor to the feature extractor. This transmission is the most relevant point for an attack on the integrity and authenticity of the acquired iris images, where the iris image inserted during the attack could be acquired with another sensor off-site, even without the knowledge of a genuine user, or be a manipulated image to spoof the biometric recognition system.

In large-scale biometric system various sensors from different manufacturers and models are deployed and the interoperability is often affected by specifics of each sensor, such as the acquisition technique or in-sensor image processing. Selective processing of the iris images helps to improve the interoperability by applying a sensor tailored biometric tool chain. Therefore information about the sensor model is required, which can be deduced from the iris images directly utilising image forensic methods.

This work evaluates the feasibility of deducing sensor information at model and unit level, i.e. the sensor an image is captured with, from the iris/ocular image using PRNU and image texture based methods. Our approach differs from existing literature in the following ways: (a) we consider both, a PRNU and an image texture based (IT) approach and analyse their strengths and weaknesses; (b) we include a larger number of iris datasets/sensors (19 datasets) and additional image forensic benchmark dataset; (c) we evaluate a fusion of the PRNU and the IT based approach to overcome the weaknesses of each single approach; (d) we consider different training set sizes down to 1 training image and different patch sizes for extracting the PRNU and the image texture features are compared; (e) we discuss the applicability of each approach as a mean of insertion attack prevention and in the context of selective processing.

The rest of the paper is organised as follows: Related work is summarised in Section 2. Section 3 describes the two different classification approaches. The experimental setup including the examined iris data sets is listed in Section 4. The results are illustrated and an application specific discussion is given in Section 5. Finally Section 6 concludes this paper.

2. Related Work

To determine an image's origin on unit level several approaches have been proposed exploiting hardware and software related artifacts. The PRNU is an intrinsic property of all digital imaging sensors due to slight variations among individual pixels in their sensitivity to incoming illumination. Consequently, every sensor casts a unique, weak, noise-like pattern onto every image it takes. This pattern, which can be regarded as a "sensor fingerprint", is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression, filtering or white-balancing. A sensor's fingerprint can be estimated from several images taken by the sensor and later detected in a given image to establish image origin and integrity.

Novel sensor identification approaches are based on Deep Convolutional Neural Networks (CNN). Tuama *et al.* [33] proposed to extract the noise residuals with a high-pass filter and classify the images using a CNN. However, this approach relies on a large number of images for the CNN training step, which limits the application scenarios for these approaches.

In the context of biometric systems security, the PRNU fingerprint of a sensor can be utilised to ensure the integrity and authenticity of images acquired with a biometric sensor. Höller *et al.* [34] propose a suitable passive approach to secure the transmission channel between the sensor and the feature extractor, making use of sensor fingerprints based on a sensor's PRNU [4].

To ensure the authenticity of the biometric sensor, first the discriminative power of the biometric sensors has to be evaluated, as it has been done in [3] and [34] using the PRNU. The results from Höller *et al.* [34], where the discriminative power of five iris sensors from the *CASIA-Iris V4* database [25] has been evaluated, show high variations. Other work by Kalka *et al.* [16] regarding the differentiability of iris sensor showed varying results. Some possible explanations are given in [16] and [34] and include highly correlated data of biometric datasets, saturated pixels and the use of multiple sensors of the same model. An additional caveat for the PRNU extraction is the image content. Since the PRNU covers the high frequency components of an image, it is contaminated with other high frequency components within the images, such as edges.

Banerjee and Ross [2] evaluated multiple PRNU estimation schemes for identifying sensors from iris images. They used 12 different datasets, 4 PRNU extraction methods and investigated dataset specific artefacts as well as the effect of a photometric transformation. They were able to identify the sensor for a majority of the datasets.

In the context of selective image processing, where it is sufficient to determine the sensor model (i.e. iris dataset), El-Naggar and Ross [11] proposed a passive approach tailored to iris recognition. At first the ocular image is segmented to get the iris region, then the iris texture is unwrapped, followed by a normalisation step to get a normalised iris image. Out of the inner half of this normalised iris image a feature vector containing statistical and Gabor features is extracted and then classified using a 3-layer artificial neural network. They were able to achieve accuracies of $80 - 85\%$.

Marra *et al.* [22] propose a CNN-based technique including transfer learning to identify the iris sensor model from iris images. They map the features extracted from images captured by one sensor to images captured by a different one. They investigated 9 different sensor models. They achieve promising results, enabling a model-adaptive preprocessing of the iris images to obtain seamless sensor interoperability.

To overcome problems in cross-sensor matching in large-scale iris recognition systems Arora *et al.* [1] developed an iris camera classification-based preprocessing framework. Using the output of their statistical image-feature based camera classification they apply a device-specific iris image enhancement leading to a significant improvement in recognition accuracy.

3. Classification Techniques

In this section we present two different techniques each allowing to infer which dataset an iris image originates from. The first technique, called PRNU based Sensor Identification (PSI), achieves this by utilising non unique artefacts embedded in the images. The second technique, Image Texture Classification (ITC), makes use of image texture information and its inherent features. Both techniques are presented in detail in the following.

3.1. PRNU based Sensor Identification (PSI)

A digital image sensor consists of lots of small photosensitive detectors, commonly known as pixels. Due to imperfections in the manufacturing and the inhomogeneity of the manufacturing material, silicon, the efficiency of each pixel in converting photons to electrons varies slightly. This slight variation is commonly known as photo-response non-uniformity (PRNU). The extraction of the PRNU noise residuals is performed as indicated by Fridrich in [13]. For each image I the noise residual W_I is estimated:

$$W_I = I - F(I) \tag{1}$$

where F is a denoising function filtering out the sensor pattern noise. Different denoising filters have been used for the extraction of the PRNU noise residual [7, 13, 24].

The extracted PRNU noise residual is then normalised in respect to the L_2-norm because its embedding strength is varying between different sensors as explained by [34].

The PRNU fingerprint $\hat{K}$ of a sensor, which isolates the systematic components and suppresses random noise, is then estimated using a maximum likelihood estimator for images I_i with $i = 1...N$.

$$\hat{K} = \frac{\sum_{i=1}^{N} W_{I_i} I_i}{\sum_{i=1}^{N} (I_i)^2} \tag{2}$$

To determine if an image has been acquired with a specific sensor, the presence of a sensor's PRNU fingerprint in the questioned image has to be detected. Since images acquired with iris sensors are usually not geometrically transformed, this can be done by means of calculating the normalised Cross Correlation (NCC) between between a PRNU noise residual of an Image J and a PRNU fingerprint weighted by the image content of J.

Furthermore, different PRNU enhancement techniques have been applied to the noise residuals and PRNU fingerprints in order to suppress undesired artifacts [19, 20].

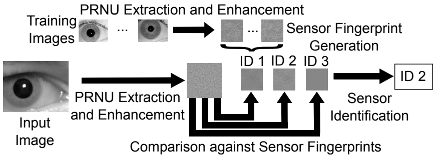

PRNU Extraction and Enhancement
Training Images
...
...
Sensor Fingerprint Generation
ID 1
ID 2
ID 3
ID 2
Input Image
PRNU Extraction and Enhancement
Sensor Identification
Comparison against Sensor Fingerprints

Figure 1. PRNU noise residual extraction and identification of corresponding sensor.

3.2. Image Texture Classification (ITC)

The ITC approach is SVM based, thus a training phase is needed, similar to generating a PRNU fingerprint for the PSI approach. The input are the iris/ocular images and the output is a prediction of the iris sensor used to capture the image or the dataset where the input iris image belongs to, respectively. In the following the three feature extraction methods, namely DenseSIFT, DMD and LBP are briefly explained. Then the classification approach using a GMM, Fisher Vector encoding and an SVM classifier is described.

3.2.1 Feature Extraction

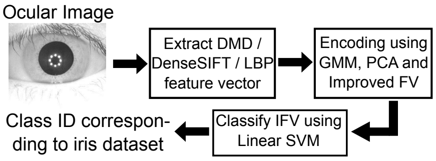

Ocular Image
Extract DMD / DenseSIFT / LBP feature vector
Encoding using GMM, PCA and Improved FV
Classify IFV using Linear SVM
Class ID corresponding to iris dataset

Figure 2. Flowchart of the Image Texture Classification (ITC) approach.

DSIFT: Fei-Fei et al. [12] proposed to use the local SIFT descriptors, a general purpose feature extraction technique used in object recognition [21], at multiple scales on a predefined grid defined across the whole image instead of localising their positions according to scale space extrema.

DMD: Dense Micro-block Difference is a local feature extraction and texture classification technique proposed by Mehta and Egiazarian [23] to capture the repetitively characteristic local structure providing discriminative information.

LBP: The local binary patterns proposed by Ojala [27] observe the variations of pixels in a local neighbourhood. These variations are thresholded against the central pixel value to obtain a binary decision, which is then encoded as a scalar value. The occurrences of each scalar value for all pixels in the image are represented in a histogram, which forms the extracted feature vector.

3.2.2 Feature Encoding

We utilise the Improved Fisher Vector Encoding (IFV) scheme in the same way as in [5]. At first the respective features (DSIFT, DMD, LBP) are extracted to obtain a feature vector f. For standard Fisher Vector (FV) encoding the feature vector f is soft-quantised using a Gaussian Mixture Model (GMM) with K modes where the Gaussian covariance matrices are assumed to be diagonal. The local descriptors present in f are first decorrelated and then dimensionality reduced (optional) by PCA. The IFV now adds signed square rooting and l^2 normalisation. For more details the interested reader is referred to [5].

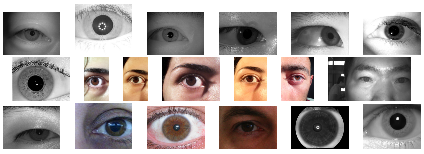


Figure 3. Sample images from different datasets.

Dataset Name	#IMG	Sensor	ILM	Resolution	CID
CASIA V2 [25]	1200	OKI IRISPASS-h	NIR	480x640	1
CASIA V3 [25]	2639	CASIA Iris camera	NIR	320x280	2
CASIA V4 [25]	20000	IrisKing IKEMB-100	NIR	640x480	3
CSIR 1 [26]	4000	EyeGuard AD100	NIR	640x480	4
CSIR 2 [26]	4000	IKEMB220	NIR	640x480	5
ICE [28]	2953	LG EOU 2200	NIR	480x640	6
IITD [18]	1120	JIRIS, JPC1000	NIR	240x320	7
MICHE S1 [9]	626	Samsung Galaxy S4 F	VL	various	8
MICHE S2 [9]	628	Samsung Galaxy S4 R	VL	various	9
MICHE S3 [9]	632	Samsung Galaxy Tab2	VL	various	10
MICHE I1 [9]	619	Apple iPhone 5 F	VL	various	11
MICHE I2 [9]	628	Apple iPhone 5 R	VL	various	12
MIR [35]	4500	Unknown Sensor	NIR	1968x1024	13
MMU2 [6]	995	Panas. BM-ET100US	NIR	320x238	14
MobBIO [32]	1640	Asus Eee Pad TE300T	VL	250x200	15
UBIRISv1 [29]	1876	Nikon E5700	VL	800x600	16
UBIRISv2 [30]	11102	Canon EOS 5D	VL	400x300	17
UPOL [10]	384	SONY DXC-950P	CF	768x576	18
UTIRIS [15]	793	ISG Lightwise LW	NIR	1000x776	19

Table 1. Attributes of iris datasets with number of images (#IMG), class ID (CID) and illumination (ILM). The illumination is either of the type near infrared (NIR), visible light (VL) or camera flash (CF).

3.2.3 Classification

A support vector machine (SVM) is used to classify the IFV encoded features. A linear kernel lead to the most promising results. The input data to the SVM (IFV encoded feature vectors) is normalised such that $K(x', x'') = 1$ which usually improves the performance. The SVM is trained using a standard non-linear SVM solver.

4. Experimental Setup

This section describes the examined datasets as well as the experimental setup.

4.1. Datasets

Table 1 summarises the most important attributes of the 19 publicly available datasets used in this work and Figure 3 shows one example image for each of the datasets. Each dataset was acquired with a distinct sensor model.

4.2. Experimental Methodology

Each dataset is randomly split into two distinct subsets, a training and a testing one. Since UPOL contains 384 images only, a 50:50 split of training and testing data results in a maximum of 192 training and 192 testing images. Datasets containing colour images are converted to greyscale. We tested different training set sizes $(1, 3, 6, 12, 24, 48, 96$ and $192)$ with a fixed test set size of 192 images for all datasets. A 5-fold cross validation is performed and the mean results of all 5 runs are the final results shown below.

All experiments are performed using different patch sizes ranging from 64×64 up to 512×512 pixels, which are cropped from the image centre. Due to the correlation based similarity measure all extracted patches must have the same size, thus the number of admissible sensors to discriminate for the PSI approach decreases with increasing patch size because of the varying image sizes among the data sets. The investigation of all 19 sensors for the PSI approach is only possible with patch sizes of 64×64 and 128×128. The ITC approach is able to handle different image sizes, hence all 19 sensors can be investigated with all patch sizes.

For the Image Texture Classification (ITC) approach the first step consists in extracting the features from the image patches using DenseSIFT, DMD and LBP. Afterwards, the features are reduced in dimensionality using a GMM and then Fisher Vector encoding is applied before they are put into a linear SVM for classification.

For the PRNU based Sensor Identification (PSI) approach the PRNU is extracted from the mentioned image patches. The extraction is performed using a variety of denoising filter and PRNU enhancement combinations, which are listed in Table 2. The interested reader is referred to the respective papers for further details on the PRNU extraction and enhancement techniques.

Name	Denoising filter	Noise residuals	Fingerprints
Li [19]	$Wavelet_{Lukas}$	Li Model 3	-
BM3D [7]	$BM3D$	-	-
FS [20]	$Wavelet_{Mihcak}$	FDR+Li	SEA

Table 2. Enhancement configurations applied to the different steps of the PRNU extraction process.

The generation of the PRNU fingerprints for the various sensors is done using the images from the "training" set. Then the NCC scores are computed for all "test" images with all generated PRNU fingerprints, where the predicted sensor (or class) is determined by means of the highest (rank one) correlation score.

Considering the score level fusion used in this work, we examined different normalisation (Minimum-Maximum, Tangens Hyperbolicus and Z-Score) and fusion schemes (Maximum, Average, Sum and Product). We tested different score combinations, from pairs of 2 scores to tuples of all 4 available scores (PSI, and the 3 ITC configurations). The Minimum-Maximum normalisation in combination with the Sum or Product fusion rule performed best across all combinations.

The following three experiments have been conducted to quantify the performance of the different techniques in discriminating between the various sensor.

Experiment 1 (EX1): Sensor Identification
The discriminability of the sensors of all iris data sets described in Table 1 using the 3 ITC (DenseSIFT, DMD, LBP) and 3 PSI (Li, BM3D, FS) configurations with 192 training and 192 test images is assessed. A patch size of 128 is used to be able to evaluate the performance for all sensors. Eventually, a score level fusion has been investigated.

Experiment 2 (EX2): Varying Patch/Training Set Sizes
Here the impact of the number of training images on the sensor identification performance of the ITC and PSI techniques is investigated. In contrast to the first experiment different training set sizes from 192 down to 1 and different patch sizes from 512 to 64 are examined. Again, a score level fusion has been investigated.

Experiment 3 (EX3): Intra-Model Sensor Identification
This experiment differs from the first two. The goal is to investigate whether the PSI and ITC techniques are able to distinguish different instances of the same sensor model. Since this is not possible with the biometric data described in Table 1, images from 6 different camera models (Casio EX-Z150, Kodak M1063, Nikon S710, Olympus MJU, Praktica DCZ 5.9 and Ricoh GX100) with 5 camera instances each have been selected from the Dresden database [14] to at least clarify this issue in general. The patch size for this experiment is 512. The training set size and test set size are set to 100 and 50, respectively, because of the low number of images available for some cameras. The discriminability of the instances has been evaluated separately for each camera model.

5. Experimental Results

In the following the results are presented and discussed. Based on the outcome of EX1 only the best performing ITC and PSI approaches have been considered for EX2 and EX3, which are: DSIFT, DMD, LBP and BM3D. The mean accuracy (mAcc) corresponds to the mean of the values of the confusion matrix diagonal. The average precision (AP) describes the area under the precision/recall curve calculated per class. The mAP is the mean over all AP values.

	DSIFT	DMD	LBP	BM3D	Li	FS	BDDF
mACC	98.78	88.51	91.96	67.82	65.92	60.20	99.48
mAP	99.51	91.23	95.05	67.93	65.26	40.72	99.86

Table 3. Mean accuracy (mACC) and mean average precision (mAP) for patch size 128 and training set size 192 for all iris datasets.

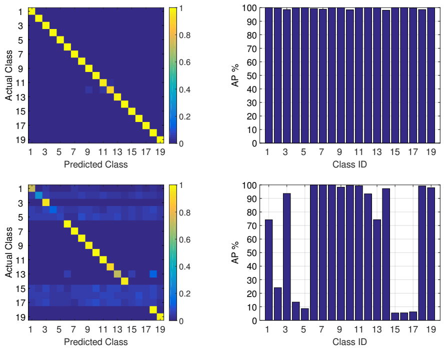


Figure 4. Confusion matrix and average precision plot for patch size 128 and training set size 192. Top: DSIFT, Bottom: BM3D.

Experiment 1 The first results listed in Table 3 are devoted to EX1. It can be seen that DSIFT performs remarkably well in distinguishing the origin of images between the various iris datasets. Figure 4 (top) confirms that DSIFT is able to determine the origin of an iris image with a very high accuracy for all of the datasets. The different PSI configurations are inferior compared to the ITC approaches. On one hand, the patch size of 128 is relatively small for a PRNU approach. On the other hand, Figure 4 (bottom) reveals that especially the classes 2, 4, 5, 15, 16 and 17 cause problems. The numbers on the axes correspond to the class IDs in Table 1. The CASIA V3 (class ID 2) dataset is suspect to contain images from multiple sensors of the same model, as already reported in literature [34, 11]. The images from the MobBIO, UBIRISv1 and UBIRISv2 datasets (classes 15, 16 and 17) have been acquired with a high resolution camera. After thorough investigation we found out that the images contained in the datasets have been cropped from different parts of the original image which causes low correlation scores for images within the same dataset. To overcome this problem these images have to be pre-aligned e.g. by using a PRNU based approach [17] or by using the peak correlation energy (PCE) measure [13]. The best score level fusion combination BDDF, which denotes the fusion of BM3D-DSIFT-DMD, improves the identification performance to a small degree.

Experiment 2 Table 4 and Figure 5 give an overview of the results for varying patch sizes and training set sizes for ITC, PSI and the fusion combination BDDF. To keep the results concise we only list some of the tested configurations. It is interesting to see that the performance of the ITC ap-

		mACC					mAP				
PS	TSS	DSIFT	DMD	LBP	BM3D	BDDF	DSIFT	DMD	LBP	BM3D	BDDF
512	192	99.92	86.09	97.96	90.27	99.96	99.98	87.17	99.12	90.23	99.99
512	24	98.90	83.07	87.50	88.89	99.18	99.49	84.61	90.16	89.14	99.59
512	3	93.16	73.59	0.00	79.63	92.69	95.19	75.56	0.00	81.08	94.86
256	192	99.64	90.70	96.67	75.03	99.87	99.92	91.80	98.22	75.52	99.97
256	24	97.96	87.90	80.04	70.70	98.30	98.93	88.10	81.97	70.75	99.16
256	3	89.29	74.58	0.00	55.38	88.31	91.85	75.62	0.00	55.30	91.29
128	192	98.78	88.51	91.96	67.82	99.48	99.52	91.23	95.05	67.93	99.68
128	24	94.78	83.76	67.75	57.48	95.49	96.48	84.36	68.31	57.00	97.13
128	3	80.26	67.09	0.00	34.23	78.79	84.13	68.03	0.00	32.37	83.67
64	192	94.95	86.86	84.07	50.93	97.69	97.13	87.59	87.77	48.21	99.01
64	24	85.57	76.28	55.05	35.28	88.32	89.13	77.27	53.31	30.34	92.03
64	3	62.91	53.17	0.00	18.68	65.75	67.78	54.36	0.00	14.83	71.18

Table 4. Results for different patch (PS) and training set sizes (TSS).

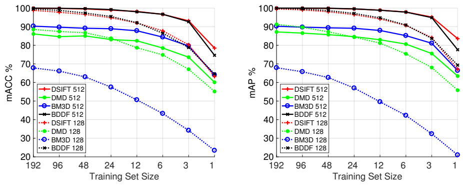


Figure 5. Results for selected patch sizes and different training set sizes.

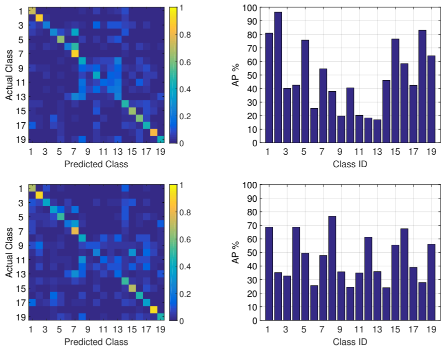


Figure 6. Confusion matrix and average precision plot for patch size 64 and training set size 1. Top: BDDF, Bottom: DSIFT.

proaches is insensitive to the training set size down to 24 images, whereas BM3D in combination with smaller patch sizes exhibits a constant performance drop towards smaller training set sizes. For larger patch sizes BM3D's performance interestingly remains almost stable down to 12 training images and its performance degrades less than the other approaches. Again fusion does not improve the overall performance, except in the case of a single training image.

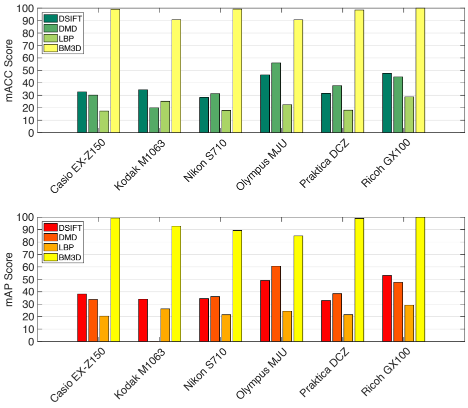


Figure 7. Results for the different camera models from the Dresden dataset with patch size 512 and training set size 100.

In Figure 6 we look at the most challenging case, patch size 64 and training set size 1, in more detail. As it can be seen in the confusion matrix and average precision plot for BDDF fusion the identification performance varies highly among the different classes resulting in an mAP of 49.45% and mACC of 46.28%. DSIFT achieves an mAP of 45.61% and mACC of 43.83% respectively. While the fusion gains accuracy for some classes (e.g. $1, 2, 15, 18$), it decreases the accuracy for other classes, leading to a slightly improved overall accuracy.

Experiment 3 This experiment reveals some interesting results regarding intra-model discrimination, which are presented in Figure 7. The BM3D approach reliably discriminates multiple instances of the same sensor model and exhibits mACC and mAP scores in the range of 82% to 100%, respectively. Despite the large patch and training set size, the ITC approaches face severe problems, with mACC and mAP scores between 0% and 60%. The ITC results strongly suggest that this approach is not useful to distinguish multiple instances of the same sensor model for arbitrary images.

5.1. Application Specific Discussion

As motivated in the introduction, identification of the image origin plays a major role for the security and performance of an iris recognition system. While it is sufficient to distinguish the origin at model level for performance enhancements, it is necessary to distinguish the origin at unit level to strengthen the security of the system.

It can be clearly seen that both, the ITC and PSI approach, are able to identify the source sensor model (i.e. iris dataset) of iris images in general. Eventually, our ITC approach outperformed the previous approach by El Naggar *et al.* [11]. However, our approach differs from the one by El Naggar *et al.*, which uses unrolled iris textures for the identification of the datasets.

The PSI approach is mostly limited by the patch size and therefore faces limited application with sensors that output low-resolution images. Pre-alignment of the images or PCE as similarity measure is necessary for the PSI approach to work properly if arbitrary cropped and resized images are present. ITC works well in distinguishing the sensor model, provided that there are sufficient training images available (more than 12). It still works for small patch sizes and especially for the classes where the PSI approach is no longer able to provide a reasonable accuracy. Consequently, the ITC approach is well suited to provide the sensor model in the context of the selective processing scenario.

The results of EX3 exposed a weakness of the ITC approach, in distinguishing arbitrary natural scene images acquired with multiple instances of the same sensor model. Hence, the ITC approach might not be the preferred solution for the insertion attack detection scenario. Following Kerckhoff's principle, i.e. assuming that an attacker knows how the whole biometric system is designed, he could simply use the same sensor model as deployed in the system to acquire a malicious image, which could then successfully bypass an ITC based attack detection system. However, as pointed out by the EX3 results, the PSI approach is able to successfully discriminate different instances of the same sensor model. Therefore, the PSI approach is able to detect such a maliciously acquired and inserted image, but its performance depends on the patch size.

Obviously, a combination of both, the ITC and PSI approach, is beneficial to overcome the individual weaknesses and improve the detection of insertion attacks. We realised this combination in form of a score level fusion. The experimental results confirmed a performance improvement.

6. Conclusion

In this paper we investigated a passive approach to deduce sensor information solely from iris images. This information is useful in forensic scenarios, e.g. for for securing an iris recognition system against insertion attacks, as well as in non-forensic ones, e.g. to enable sensor model specific selective processing of the images. Our approach is based on two different techniques, a PRNU (PSI) and a texture classification one (ITC). In addition a score level fusion of the two different techniques is investigated to further improve the performance. Our experiments include tests using different numbers of training images as well as different image patch sizes.

The results confirm that our approach is well suited to identify the source sensor model of a given iris images in all test cases. It achieves almost 100% accuracy given that the training set size and patch size are sufficiently large. It still works reasonably well even for low resolution input images. The PSI approach is able to distinguish different sensors at unit level, but requires a certain patch size. By combining ITC and PSI through score level fusion a unit-level discrimination becomes possible for a broad range of sensor configurations.

Since no biometric dataset covering several units of the same sensor model is publicly available, we aim at establishing such a dataset. Our future work will then include extended tests to shed more light at the unit-level discrimination performance of our approach as well as investigation of an open set scenario.

Overall, by identifying the image origin at model and unit level, our approach forms the basis for the application of sensor specific processing of the iris images and can be of particular interest for securing iris recognition systems.

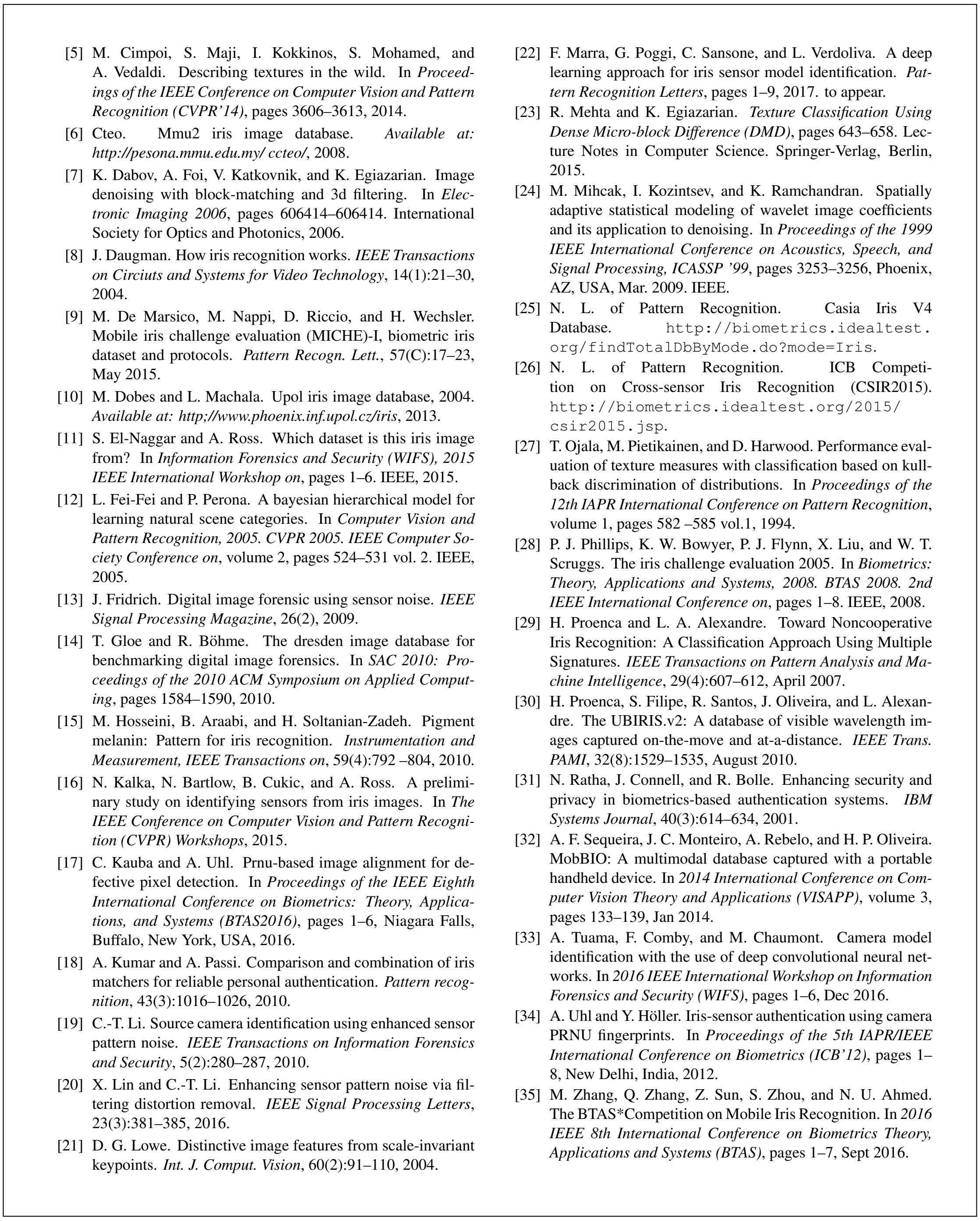
References

[1] S. S. Arora, M. Vatsa, R. Singh, and A. Jain. On iris camera interoperability. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 346–352, Sept 2012.

[2] S. Banerjee and A. Ross. From image to sensor: Comparative evaluation of multiple prnu estimation schemes for identifying sensors from nir iris images. In *Proceedings of 5th International Workshop on Biometrics and Forensics (IWBF'17)*, Coventry, UK, 2017.

[3] N. Bartlow, N. Kalka, B. Cukic, and A. Ross. Identifying sensors from fingerprint images. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 78–84, 2009.

[4] M. Chen, J. Fridrich, M. Goljan, and J. Lukas. Determining image origin and integrity using sensor noise. *IEEE Transactions on Information Security and Forensics*, 3(1):74–90, 2008.



[5] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*, pages 3606–3613, 2014.

[6] Cteo. Mmu2 iris image database. *Available at: http://pesona.mmu.edu.my/ ccteo/*, 2008.

[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising with block-matching and 3d filtering. In *Electronic Imaging 2006*, pages 606414–606414. International Society for Optics and Photonics, 2006.

[8] J. Daugman. How iris recognition works. *IEEE Transactions on Circiuts and Systems for Video Technology*, 14(1):21–30, 2004.

[9] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recogn. Lett.*, 57(C):17–23, May 2015.

[10] M. Dobes and L. Machala. Upol iris image database, 2004. *Available at: http;//www.phoenix.inf.upol.cz/iris*, 2013.

[11] S. El-Naggar and A. Ross. Which dataset is this iris image from? In *Information Forensics and Security (WIFS), 2015 IEEE International Workshop on*, pages 1–6. IEEE, 2015.

[12] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 524–531 vol. 2. IEEE, 2005.

[13] J. Fridrich. Digital image forensic using sensor noise. *IEEE Signal Processing Magazine*, 26(2), 2009.

[14] T. Gloe and R. Böhme. The dresden image database for benchmarking digital image forensics. In *SAC 2010: Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 1584–1590, 2010.

[15] M. Hosseini, B. Araabi, and H. Soltanian-Zadeh. Pigment melanin: Pattern for iris recognition. *Instrumentation and Measurement, IEEE Transactions on*, 59(4):792 –804, 2010.

[16] N. Kalka, N. Bartlow, B. Cukic, and A. Ross. A preliminary study on identifying sensors from iris images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2015.

[17] C. Kauba and A. Uhl. Prnu-based image alignment for defective pixel detection. In *Proceedings of the IEEE Eighth International Conference on Biometrics: Theory, Applications, and Systems (BTAS2016)*, pages 1–6, Niagara Falls, Buffalo, New York, USA, 2016.

[18] A. Kumar and A. Passi. Comparison and combination of iris matchers for reliable personal authentication. *Pattern recognition*, 43(3):1016–1026, 2010.

[19] C.-T. Li. Source camera identification using enhanced sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 5(2):280–287, 2010.

[20] X. Lin and C.-T. Li. Enhancing sensor pattern noise via filtering distortion removal. *IEEE Signal Processing Letters*, 23(3):381–385, 2016.

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[22] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva. A deep learning approach for iris sensor model identification. *Pattern Recognition Letters*, pages 1–9, 2017. to appear.

[23] R. Mehta and K. Egiazarian. *Texture Classification Using Dense Micro-block Difference (DMD)*, pages 643–658. Lecture Notes in Computer Science. Springer-Verlag, Berlin, 2015.

[24] M. Mihcak, I. Kozintsev, and K. Ramchandran. Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising. In *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '99*, pages 3253–3256, Phoenix, AZ, USA, Mar. 2009. IEEE.

[25] N. L. of Pattern Recognition. Casia Iris V4 Database. `http://biometrics.idealtest.org/findTotalDbByMode.do?mode=Iris`.

[26] N. L. of Pattern Recognition. ICB Competition on Cross-sensor Iris Recognition (CSIR2015). `http://biometrics.idealtest.org/2015/csir2015.jsp`.

[27] T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, volume 1, pages 582 –585 vol.1, 1994.

[28] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, and W. T. Scruggs. The iris challenge evaluation 2005. In *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, pages 1–8. IEEE, 2008.

[29] H. Proenca and L. A. Alexandre. Toward Noncooperative Iris Recognition: A Classification Approach Using Multiple Signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):607–612, April 2007.

[30] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. Alexandre. The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance. *IEEE Trans. PAMI*, 32(8):1529–1535, August 2010.

[31] N. Ratha, J. Connell, and R. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[32] A. F. Sequeira, J. C. Monteiro, A. Rebelo, and H. P. Oliveira. MobBIO: A multimodal database captured with a portable handheld device. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 3, pages 133–139, Jan 2014.

[33] A. Tuama, F. Comby, and M. Chaumont. Camera model identification with the use of deep convolutional neural networks. In *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, Dec 2016.

[34] A. Uhl and Y. Höller. Iris-sensor authentication using camera PRNU fingerprints. In *Proceedings of the 5th IAPR/IEEE International Conference on Biometrics (ICB'12)*, pages 1–8, New Delhi, India, 2012.

[35] M. Zhang, Q. Zhang, Z. Sun, S. Zhou, and N. U. Ahmed. The BTAS*Competition on Mobile Iris Recognition. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, Sept 2016.

Blind Source Camera Clustering of Criminal Case Data

Luca Debiasi*, Elisabet Leitet†, Kristin Norell‡, Theodoros Tachos† and Andreas Uhl*
*WaveLab - The Multimedia Signal Processing and Security Lab, University of Salzburg, Salzburg, Austria
†NFC - Swedish National Forensic Centre, Swedish Police Authority, Linköping, Sweden
‡Formerly NFC - Swedish National Forensic Centre, Swedish Police Authority, Linköping, Sweden
{ldebiasi,uhl}@cs.sbg.ac.at
{elisabet.leitet,theodoros.tachos}@polisen.se
kristin.norell@gmail.com

***Abstract*—This work focuses on the examination of a real word criminal case data set, consisting of still images found on a suspect's computer during a sexual abuse case investigation. Various source camera clustering algorithms, all based on the photo-response non-uniformity (PRNU), are employed to organise the images according to their source camera. The investigated data set poses many challenges to the algorithms due to the unknown origin of its images. The clustering result's quality is examined using different external and internal cluster validity indices (CVIs). Before attempting to cluster the criminal case data, the clustering algorithms and CVIs have been examined on a different data set with known ground truth, which revealed that some algorithms and CVIs are not appropriate for this scenario.**

Finally, we give some recommendations on which clustering algorithms and CVIs can be used in this scenario and discuss the problems and challenges we faced while investigating the data set.

***Index Terms*—Digital Image Forensics, Source Sensor Clustering, PRNU, Criminal Case Investigation.**

I. Introduction

In forensic case work, source camera identification using PRNU can yield important evidence for the criminal investigation. The properties of the examination makes it well suited to be used in a Bayesian evaluation scheme as a Likelihood Ratio (LR) calculation [1]. Typical criminal cases, where such examinations can be of use, include fraud, sexual child abuse, rape and assault. Often, the perpetrators have an urge of documenting their criminal actions, and the imagery is often captured by mobile phones readily at hand. In the ideal case, the suspect's camera is available for collection of all necessary reference data to conduct the examination. A detailed EXIF data analysis is always complementing the PRNU examination and is included in the evaluation. When signs of alteration are found in the EXIF data, the weight of the PRNU examination will be lower and the LR value approaches unity.

However, in cases where reliable reference data can not be obtained, it can be useful to organise the images confiscated on the suspect's computer by their source camera instead. These images should have general properties fitting that of the questioned imagery, and preferably also a connection to the suspect (e.g. family album). For this scenario, a prior screening of the data based on the image origin or source camera could be very useful. Source camera clustering based on the camera's PRNU offers an intuitive solution to this problem by associating images that have been captured with the same device. Such information could be important to identify the number of victims in grooming cases and to find more images taken with the same webcam (victim), or to evaluate the number of perpetrators in sexual child abuse cases.

In the source camera clustering scenario, however, the investigator is usually confronted with a large set of images from unknown source(s). The goal is to group all images according to the source camera, where the number of cameras as well as the distribution of the images among them is unknown. In this case it is usually not possible for the investigator to acquire additional data because the source cameras might not be available. Several classical clustering techniques have been proposed in literature to solve this problem [2–9].

As already mentioned, the source camera clustering problem is solved by partitioning the data set under investigation using a clustering algorithm. According to Wang *et al.* [10] the term *cluster validity assessment* describes the process of evaluating the clustering result. This evaluation is based on two criteria, which are used to determine the "optimal" clustering solution:

- *Compactness*: The members of each cluster should be as close to each other as possible.
- *Separation*: The clusters themselves should be widely separated.

The partition that best fits the underlying data can be considered as the "optimal" clustering solution. Several clustering validity indices (CVIs) have been proposed in literature, which can be divided into external and internal indices (or criteria) [11]: An external index is a measure of agreement between two clusterings where the reference clustering is known a priori, and the second results from a clustering procedure. Internal indices are used to measure the quality of a clustering structure without external information. For external indices the results of a clustering algorithm based on a known cluster structure of a data set (or cluster labels) are evaluated, while for internal indices the results are evaluated using quantities and features

This work is partially funded by the Austrian Science Fund (FWF) under Project No. P26630 and partially supported by a COST 1106 Short Term Scientific Mission (STSM).

TABLE I: Properties of images in criminal case data set: number of examinable images for different image sizes, number of those images containing EXIF metadata and number of different camera models in EXIF data.

Image Size	Exam. Imgs.	Imgs. EXIF (%)	# Cameras
$\geq 256 \times 256$	3078	2097 (~68%)	60
$\geq 512 \times 512$	1961	1006 (~51%)	47
$\geq 1024 \times 1024$	851	765 (~90%)	35

inherent in the data set. The optimal number of clusters is usually determined based on an internal validity index.

The main contribution of this paper is to show the challenges of source camera clustering in a real world application and to give an incentive for future research in this field. The paper is organised as follows: Section II explains the motivation for this work and some further details about the criminal case, Section III describes the examined criminal case data set, Section IV describes the experimental set up, Section V illustrates the results of our experiments and the challenges faced during the investigation, while Section VI concludes the paper.

II. Motivation

The study presented in this paper is based on a criminal case investigated by the Swedish Police Authority. The Swedish National Forensic Centre (NFC) was consulted by the investigators regarding methods of victim identification in large collections of images.

According to the investigators, an offender had been communicating and interacting with young adolescents through an internet based communication application transmitting both video and audio. The investigators also had information that still images had been sent from the various victims to the offender's computer and observed that the confiscated computer of a suspect contained a large amount of still images. Due to the amount of data, the investigators requested a solution for automatically processing this large collection of images, with the aim of finding potential victims within it. The large number of images made manual processing of each image unfeasible.

As part of this, NFC suggested that a clustering approach could perhaps be performed by examining the PRNU of images found on the confiscated computer. If images could be organised by image source, the search for compromising material depicting the victims could be performed more efficiently on a per-source basis. A methodology of clustering images from unknown recording units based on PRNU was not in use at NFC at that time. Due to time constraints, the PRNU approach was abandoned in this specific instance. However, the development of such a method for use in future investigations has led to the study presented here.

III. Criminal Case Data Set

The study presented in this paper is performed on digital still images extracted from the criminal case presented in the previous Section II. Still images, both allocated and unallocated in the file system, have been extracted from the investigated computer. Allocated data files are accessible and readable by means of the file system on the digital storage device, whereas unallocated data files are not. The unallocated still images have been recovered using both commercial and non-commercial forensic tools. These images are filtered based on their uniqueness (calculated hash value) and file size f_z, being in the range of 10 KB $\leq f_z \leq$10 MB. The final data set contains 3078 images.

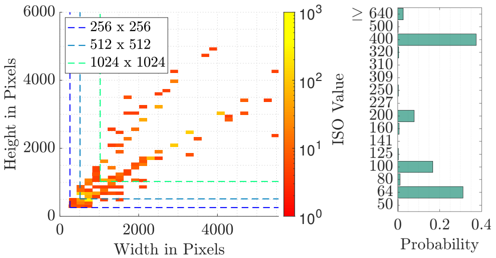


Fig. 1: Distribution of image resolutions and ISO sensitivity of images in the criminal case data set.

Figure 1 depicts the distribution of the images' resolutions and the images usable for different PRNU sizes, i.e. the size of the extracted PRNU patch. From the graphs it is noticeable that the number of the examinable images decreases as the extracted PRNU's size increases, because only images with a size larger than the PRNU size can be examined. Furthermore, it shows a histogram of the different ISO sensitivities used to acquire the images.

Additional metadata information being stored in the EXIF data is extracted and analysed. The EXIF metadata may contain camera model names, suggesting that some of the images might originate from the same camera model/unit. Table I lists the number of examinable images and different camera models retrieved from the EXIF data for each PRNU size.

IV. Experimental Setup

The goal of this work is to cluster images from potentially multiple sensors in the data described in Section III, which was found on a computer during a criminal case. The investigation has been performed by extracting the PRNU with different sizes from the image center: 256×256, 512×512 and 1024×1024 pixels. This enables us to compare the PRNU of images with different image sizes, which are mentioned in Section III. The number of images available for the investigation decreases with increasing PRNU size, which poses a trade-off between the two. The PRNU extraction and calculation of the PRNU fingerprints have been performed as proposed by Fridrich in [12], but the Block-matching and 3D filtering (BM3D) filter proposed by Dabov *et al.* [13] is used instead of the proposed wavelet-based denoising filter. According to [14, 15], BM3D is reported to yield a more consistent PRNU extraction on large data sets compared to other denoising filters.

Four different source camera clustering techniques, based on three distinct clustering principles, are investigated in this work:

- **Agglomerative clustering**: Blind Camera Fingerprinting and Image Clustering (BCF)[5]
- **Hierarchical clustering**: Unsupervised Clustering of Digital Images (UCDI)[2], Fast Image Clustering (FICL)[3]
- **Spectral clustering**: Multi-Class Spectral Clustering (MCSC) [16]

The outcome of all algorithms is a list of clusters with associated images. More details on the algorithms can be found in the corresponding papers. The clustering results of the various source camera clustering algorithms are evaluated in form of a cluster validity assessment, as described in Section I. The internal CVIs used in this work, all computed using the CVAP toolbox [10], are:

- **Davies-Bouldin Index (DBI)** [17]: Reflects the average similarity between a cluster and its most similar one.
- **Silhouette Index (SI)** [18]: Index measuring the compactness and separation of clusters.
- **Calinski-Harabasz index (CHI)** [19]: Measures between-cluster isolation and within-cluster coherence.
- **Dunn Index (DI)** [20]: Index that maximises inter-cluster distances, while minimising intra-cluster ones.

For DBI, smaller values indicate compact and well-separated clusters, while for CHI and DI this is indicated by larger values. For SI, negative values indicate an incorrect clustering, values around 0 overlapping clusters and positive values a dense clustering with high compactness and separation. Furthermore, the following external CVIs have been computed using the Scikit-learn toolbox (https://scikit-learn.org):

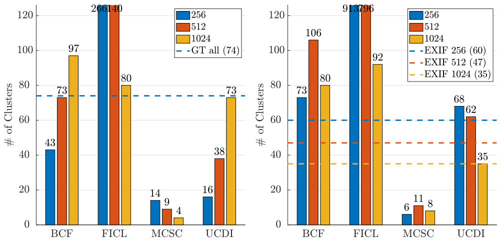
- **Homogeneity (HOM)** [21]: Measures if only members of the same class are assigned to a cluster.
- **Completeness (COM)** [21]: Measures if all members of the same class are assigned to the same cluster.
- **Adjusted Mutual Information (AMI)** [22]: Measures the agreements of two clusterings ignoring permutations and normalised against chance.
- **Adjusted Rand Index (ARI)** [23]: Measures the similarity of two clusterings with adjustment for chance.

For all external indices, higher values (closer to 1) indicate better results. AMI and ARI furthermore are adjusted against chance, which means that they have a score of 0 when the result could also be obtained by chance alone. For further details on the various indices, the reader is referred to the corresponding papers.

In order to assess the general performance of the used source camera clustering algorithms and validity and reliability of the CVIs, a source camera clustering has been performed on a subset of the Dresden Image Database [24] first, where 30 images have been randomly selected for each of the 74 distinct cameras.

V. Results and Discussion

As described in the previous section, two different experiments have been conducted in this work: First, the clustering algorithms and CVIs are evaluated on a subset of the Dresden Image Database with known ground truth. With the knowledge gained from the first experiment, the same algorithms and metrics are applied to the criminal case data. The results of both experiments are presented in the remainder of this section together with a discussion of the results.



(a) Dresden Image DB

(b) Criminal Case

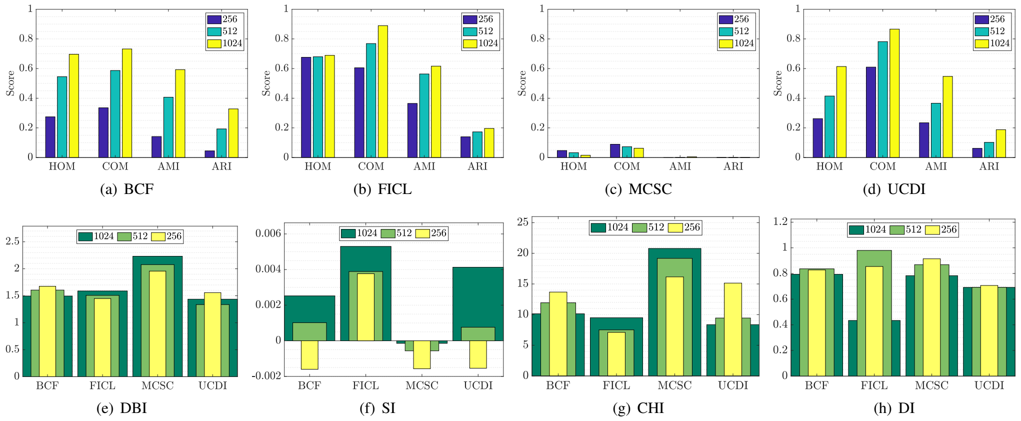
Fig. 2: Number of obtained clusters for the examined data sets.

A. Dresden Image DB

To begin with, the number of resulting clusters obtained from applying the various source camera clustering algorithms is illustrated in Figure 2(a). It can be observed, that a larger PRNU size leads to an increase of clusters for BCF and UCDI, while a decrease of the cluster number can be observed for FIC and MCSC. When looking at the ground truth number of 74 clusters, BCF, FICL and UCDI come very close to it with PRNU sizes of 512×512 and 1024×1024, while MCSC yields very low cluster numbers for all PRNU sizes. It can also be observed that FICL produces a very high number of clusters compared to all other algorithms.

Obviously, the quality of the clustering outcome does not rely on the number of resulting clusters alone. Thus, the results of the external CVI are presented in Figure 3(a) to 3(d). As expected from the number of clusters, MCSC shows the lowest metric scores of all algorithms, making this algorithm unable to cluster the data set properly. Since AMI and ARI are almost equal to 0, the resulting cluster structure is almost equivalent to a random assignment. BCF and UCDI show a very similar behaviour: the larger the PRNU size, the better the metric scores. With the largest PRNU size of 1024×1024 pixels, good results can be achieved when looking at all external CVIs, even the ones adjusted for chance. The overall best results are achieved by FICL, which has the highest scores of all investigated algorithms among all external clustering indices. In particular, the stable HOM scores across all PRNU sizes and the growing COM scores with larger PRNU sizes are noteworthy.

Figure 3(e) to 3(h) illustrates the internal CVIs' results. At first glance, DBI and SI seem to reflect the external CVIs' results, while CHI and DI do not. Furthermore, CHI seems to be biased against a low number of clusters because MCSC for all PRNU sizes and UCDI for 256×256 yield high scores. DI indicates that the performance of MCSC is on par or even better than other algorithms, which contradicts the previous external CVIs' results. Hence, CHI and DI do not seem to



(a) BCF (b) FICL (c) MCSC (d) UCDI

(e) DBI (f) SI (g) CHI (h) DI

Fig. 3: External (a-d) and internal (e-h) CVI scores for the Dresden Image DB experiment.

be able to reliably assess the clustering performance in this scenario. The DBI scores are very similar for the different PRNU sizes, though rather large differences can be observed in the previous external CVIs results. Nonetheless, the general performance trends of the various algorithms are resembled in the scores. Contrary to all other internal CVIs, the SI scores have the highest consensus with the external CVIs regarding the algorithm's performance differences for the various PRNU sizes as well as the relation performance differences among the different algorithms themselves. Thus, SI can be considered as the most trustworthy internal CVI in this case.

B. Criminal Case Data

With previous results on ground truth data in mind, we now focus on the criminal case data presented in Section III. It has to be noted, that for larger PRNU sizes less images can be examined (illustrated in Table I). As it can be observed in Figure 2(b), the various algorithms show a very similar behaviour to the clustering of the Dresden Image DB. FICL exhibits a very high amount of clusters, while MCSC exhibits a very low amount. The number of clusters of UCDI is very close to the number of models, while BCF's one is above it.

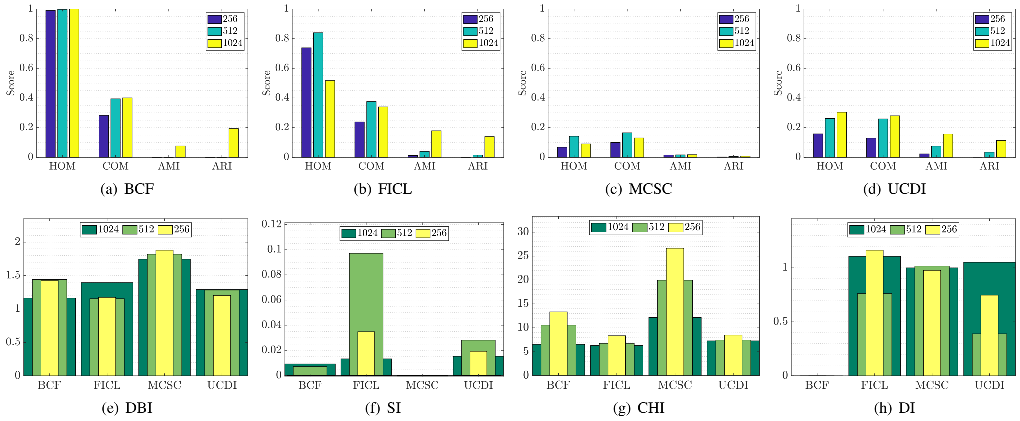
The results of the external CVIs are illustrated in Figure 4(a) to 4(d). In order to evaluate the external CVIs for this data set, some assumptions had to be made: The reference clustering structure was generated with the EXIF data's camera model information, where images without metadata have been excluded. In general, the external CVI scores of all algorithms are significantly lower than the scores obtained on the Dresden Image DB. Only HOM is very high, especially for BCF which did not exhibit such high scores for ground truth data. Considering the overall results, only FICL and BCF seem to produce reasonable results. However, these results have to be interpreted with caution, because the EXIF information might have been manipulated and the reference clustering is based on camera models and not unit level. Though, the number of different camera models in the EXIF metadata could be seen as lower bound for the expected number of clusters.

For the evaluation of the internal CVIs, all examinable images are considered again for computing the internal CVI scores, which are illustrated in Figure 4(e) to 4(h). CHI and DI again show unintuitive results, which contradict all other internal and external CVI results. DBI shows similar scores to the clustering of the Dresden DB and attributes similar performance to all clustering algorithms except MCSC, while SI shows much higher performance gaps between the various algorithms. FICL again yields the highest but highly variable scores in this scenario, while a lower but consistent performance is achieved by UCDI.

C. Discussion and Recommendations

The clustering of the criminal case data set poses many challenges. It contains images from an unknown number of different cameras taken under unknown acquisition conditions and the images might have been subject to unknown post-processings, such as cropping, scaling, rotation, contrast enhancement and other transformations. Datasets used in literature mostly contain images evenly distributed among different cameras, which are acquired under controlled conditions using the base ISO sensitivity of the cameras. In reality, however, images might cover a wide range of different ISO sensitivities, as shown in Figure 1. To the author's best knowledge, there is almost no literature which investigates and, more importantly, proposes a solution to these challenges. Regarding scaling and cropping, a brute force parameter search [25] and the use of a computationally expensive filter (MACE-MRH) have been proposed [26]. Both of these are not feasible for a clustering scenario, since they must be recomputed for every image comparison. Furthermore, the effects of denoising, recompression and demosaicing on the PRNU have been investigated in [27].

The data set furthermore contains images with different resolutions as illustrated in Figure 1. It is well known, that



(a) BCF (b) FICL (c) MCSC (d) UCDI

(e) DBI (f) SI (g) CHI (h) DI

Fig. 4: External (a-d) and internal (e-h) CVI scores for the criminal case data experiment.

a larger PRNU size leads to more reliable results, which we confirmed in the first experiment investigating the Dresden Image DB. However, when working with the criminal case data set investigated in this work a forensic expert has to deal with the trade-off between image size and number of images available for investigation. The decision is made even more difficult due to the fact that most images are of smaller size and would not be examinable when choosing a larger PRNU size.

The examined CVIs are also shown to not be very consistent, especially the internal CVIs show contradicting results. Because only internal CVIs can be used in a scenario with no ground truth data, as in the case of the criminal case data set, this leaves the selection of a reliable CVI an open question. Our results suggest that the Silhouette Index (SI) might be the most reliable index among the examined ones.

An alternative approach would be to use the EXIF metadata to generate a reference clustering on model level and then employ external CVIs to evaluate the resulting clustering, as described in the previous section. However, this metadata could potentially be manipulated and therefore not trustworthy. Furthermore, images with missing EXIF information cannot be examined with this approach. For this approach, we recommend to use either the Adjusted Mutual Information (AMI) or Adjusted Rand Index (ARI) to evaluate the clustering result, due to them being adjusted for chance. This property is valuable, especially when the number of clusters is expected to be high compared to the number of investigated images.

Regarding the examined clustering algorithms, FICL shows the most consistent performance, followed by BCF and UCDI. We cannot recommend the use of MCSC in this scenario because of the obtained results. The selection of the clustering algorithm seems to be less important with increasing PRNU size. For the scenario dealt with in this work, we recommend FICL for the clustering, since the source camera clustering would mainly be used for screening purposes, as described in Section II, where the higher number of clusters compared to the other algorithms is not a substantial issue.

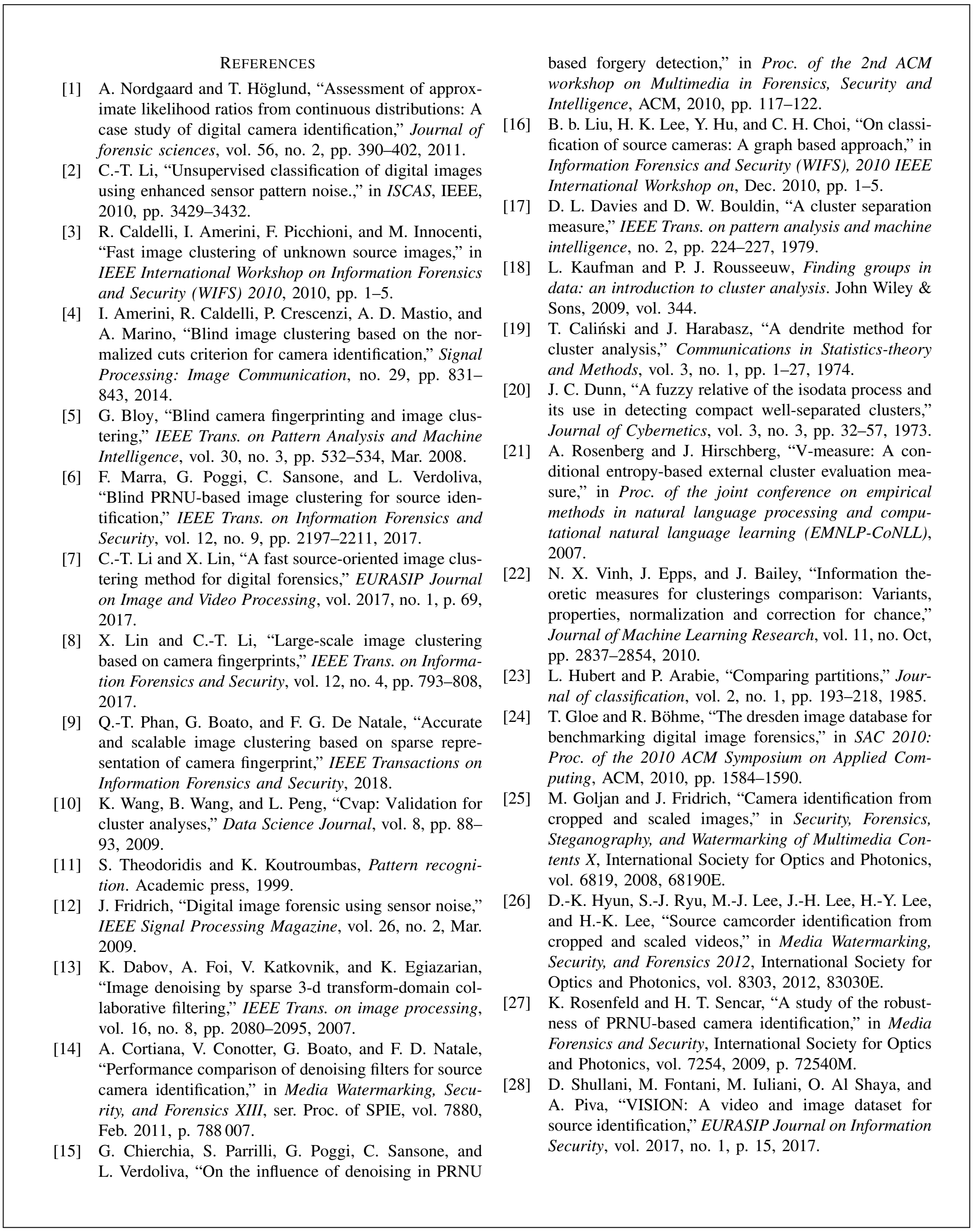
In future work, we plan to investigate more recent clustering algorithms [4, 6–9] as well as making use of the VISION dataset [28] for clustering and CVI performance evaluation.

VI. Conclusion

The main aspect of this work is to examine a data set comprised of a large amount of still images found on a suspect's confiscated computer during a criminal investigation. The data is examined by employing different PRNU-based source camera clustering algorithms, in order to organise the images by their source camera(s). Thereafter, a quantitative analysis of the clustering outcome is conducted by means of different external and internal cluster validity indices (CVIs).

Before examining the criminal data set, we need to assess the reliability and integrity of the clustering algorithms and CVIs. This assessment is performed on a subset of the Dresden Image Database, which enabled us to reveal the inability of certain algorithms and CVIs to properly cluster and quantify the output of this data with known ground truth. The knowledge gained from this preliminary analysis enabled us to better understand the contradicting results obtained when examining the criminal case data. Finally, we gave some recommendations on how to handle this kind of scenario. This challenging data set left many open questions and issues for future work, especially regarding the robustness of PRNU-based algorithms in regard to real world data and how the quality of a clustering can be reliably assessed.

Eventually, a robust and reliable source camera clustering approach could be used to build a database holding PRNU signatures of confiscated images of illicit content. If consistently updated, such a database could reveal potential connections and provide leads for further investigation.



References

[1] A. Nordgaard and T. Höglund, “Assessment of approximate likelihood ratios from continuous distributions: A case study of digital camera identification,” *Journal of forensic sciences*, vol. 56, no. 2, pp. 390–402, 2011.

[2] C.-T. Li, “Unsupervised classification of digital images using enhanced sensor pattern noise.,” in *ISCAS*, IEEE, 2010, pp. 3429–3432.

[3] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti, “Fast image clustering of unknown source images,” in *IEEE International Workshop on Information Forensics and Security (WIFS) 2010*, 2010, pp. 1–5.

[4] I. Amerini, R. Caldelli, P. Crescenzi, A. D. Mastio, and A. Marino, “Blind image clustering based on the normalized cuts criterion for camera identification,” *Signal Processing: Image Communication*, no. 29, pp. 831–843, 2014.

[5] G. Bloy, “Blind camera fingerprinting and image clustering,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 532–534, Mar. 2008.

[6] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, “Blind PRNU-based image clustering for source identification,” *IEEE Trans. on Information Forensics and Security*, vol. 12, no. 9, pp. 2197–2211, 2017.

[7] C.-T. Li and X. Lin, “A fast source-oriented image clustering method for digital forensics,” *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, p. 69, 2017.

[8] X. Lin and C.-T. Li, “Large-scale image clustering based on camera fingerprints,” *IEEE Trans. on Information Forensics and Security*, vol. 12, no. 4, pp. 793–808, 2017.

[9] Q.-T. Phan, G. Boato, and F. G. De Natale, “Accurate and scalable image clustering based on sparse representation of camera fingerprint,” *IEEE Transactions on Information Forensics and Security*, 2018.

[10] K. Wang, B. Wang, and L. Peng, “Cvap: Validation for cluster analyses,” *Data Science Journal*, vol. 8, pp. 88–93, 2009.

[11] S. Theodoridis and K. Koutroumbas, *Pattern recognition*. Academic press, 1999.

[12] J. Fridrich, “Digital image forensic using sensor noise,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, Mar. 2009.

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” *IEEE Trans. on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[14] A. Cortiana, V. Conotter, G. Boato, and F. D. Natale, “Performance comparison of denoising filters for source camera identification,” in *Media Watermarking, Security, and Forensics XIII*, ser. Proc. of SPIE, vol. 7880, Feb. 2011, p. 788 007.

[15] G. Chierchia, S. Parrilli, G. Poggi, C. Sansone, and L. Verdoliva, “On the influence of denoising in PRNU based forgery detection,” in *Proc. of the 2nd ACM workshop on Multimedia in Forensics, Security and Intelligence*, ACM, 2010, pp. 117–122.

[16] B. b. Liu, H. K. Lee, Y. Hu, and C. H. Choi, “On classification of source cameras: A graph based approach,” in *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*, Dec. 2010, pp. 1–5.

[17] D. L. Davies and D. W. Bouldin, “A cluster separation measure,” *IEEE Trans. on pattern analysis and machine intelligence*, no. 2, pp. 224–227, 1979.

[18] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009, vol. 344.

[19] T. Caliński and J. Harabasz, “A dendrite method for cluster analysis,” *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.

[20] J. C. Dunn, “A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters,” *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[21] A. Rosenberg and J. Hirschberg, “V-measure: A conditional entropy-based external cluster evaluation measure,” in *Proc. of the joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 2007.

[22] N. X. Vinh, J. Epps, and J. Bailey, “Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance,” *Journal of Machine Learning Research*, vol. 11, no. Oct, pp. 2837–2854, 2010.

[23] L. Hubert and P. Arabie, “Comparing partitions,” *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.

[24] T. Gloe and R. Böhme, “The dresden image database for benchmarking digital image forensics,” in *SAC 2010: Proc. of the 2010 ACM Symposium on Applied Computing*, ACM, 2010, pp. 1584–1590.

[25] M. Goljan and J. Fridrich, “Camera identification from cropped and scaled images,” in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, International Society for Optics and Photonics, vol. 6819, 2008, 68190E.

[26] D.-K. Hyun, S.-J. Ryu, M.-J. Lee, J.-H. Lee, H.-Y. Lee, and H.-K. Lee, “Source camcorder identification from cropped and scaled videos,” in *Media Watermarking, Security, and Forensics 2012*, International Society for Optics and Photonics, vol. 8303, 2012, 83030E.

[27] K. Rosenfeld and H. T. Sencar, “A study of the robustness of PRNU-based camera identification,” in *Media Forensics and Security*, International Society for Optics and Photonics, vol. 7254, 2009, p. 72540M.

[28] D. Shullani, M. Fontani, M. Iuliani, O. Al Shaya, and A. Piva, “VISION: A video and image dataset for source identification,” *EURASIP Journal on Information Security*, vol. 2017, no. 1, p. 15, 2017.

PRNU-based Detection of Morphed Face Images

Luca Debiasi*, Ulrich Scherhag†, Christian Rathgeb†, Andreas Uhl* and Christoph Busch†
*WaveLab – The Multimedia Signal Processing and Security Lab, Universität Salzburg, Austria
†da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
{ldebiasi,uhl}@cosy.sbg.ac.at
{ulrich.scherhag,christian.rathgeb,christoph.busch}@h-da.de

***Abstract*—In the recent past, face recognition systems have been found to be highly vulnerable to attacks based on morphed biometric samples. Such attacks pose a severe security threat to biometric recognition systems across various applications. Apart from some algorithms, which have been reported to reveal practical detection performance on small in-house datasets, approaches to effectively detect morphed face images of high quality have remained elusive. In this paper, we propose a morph detection algorithm based on an analysis of photo response non-uniformity (PRNU). It is based on a spectral analysis of the variations within the PRNU caused by the morphing process. On a comprehensive database of 961 bona fide and 2,414 morphed face images practical performance in terms of detection equal error rate (D-EER) is achieved. Additionally, the robustness of the proposed morph detection algorithm towards different post-processing procedures, e.g. histogram equalization or sharpening, is assessed.**

I. Introduction

Automated face recognition represents a longstanding field of research and a variety of methods have been proposed over the past three decades [1], [2]. Generic face recognition systems comprise four major modules: face detection, face alignment, feature extraction, and comparison, where the latter two are generally conceded as key modules. The potentially high intra-class variability within human faces across time represents a main challenge in face recognition systems. Hence, in order to achieve acceptable False Non-Match Rates (FNMRs) deployments of face recognition systems are operated at rather high False Match Rates (FMRs) [3].

In past years, researchers have pointed out diverse potential vulnerabilities of biometric recognition systems [4]. In particular, face recognition systems have been found to be vulnerable to presentation attacks [5]. Presentation attacks refer to a presentation of an attack instrument (e.g. print outs or electronic displays) to the biometric capture device with the goal of interfering with the operation of the biometric recognition system [6]. More recently, attacks on face recognition systems based on morphed biometric images have been presented [7], [8], which represent a presentation attack at the time of enrolment. Morphing techniques can be used to create artificial biometric samples, which resemble the biometric information of two (or more) individuals in image and feature domain. If morphed biometric images are infiltrated to a biometric recognition system during enrolment the subjects contributing to the morphed image will both (or all) be successfully verified against that single enrolled template. Hence, the unique link between individuals and their biometric reference data is not warranted. Fig. 1 shows an example of morphing two facial images.

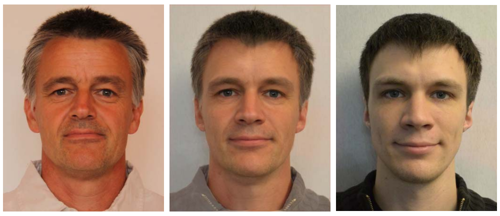

(a) Subject 1 (b) Morph (c) Subject 2

Fig. 1: Examples for bona fide and morphed face images

Attacks based on morphed biometric samples were first introduced by Ferrara *et al.* [7]. Motivated by security gaps in the issuance process of electronic travel documents, the authors showed that commercial face recognition software tools are highly vulnerable to such attacks, i.e. different images of either subject are successfully matched against the morphed image. In their experiments, decision thresholds yielding a FMR of 0.1% have been used, according to the guidelines provided by the European Agency for the Management of Operational Cooperation at the External Borders (FRONTEX) [3]. In a further study, the authors show that morphed face images are realistic enough to fool human examiners [9]. Scherhag *et al.* [8] reported moderate detection performance for benchmarking several general purpose texture descriptors used in conjunction with machine learning techniques to detect morphed face images. With respect to the above attack scenario, it is stressed that a detection of morphed face images becomes even more challenging if images are printed and scanned. Hildebrandt *et al.* [10] suggest to employ generic image forgery detection techniques, in particular multi-compression anomaly detection, to reliably detect morphed facial images. Kraetzer *et al.* [11] evaluate the feasibility of detecting facial morphs with keypoint descriptors and edge operators. The benefits of deep neural networks for detecting morphed images has been recently investigated by Ramachandra *et al.* [12].

Gomez-Barrero *et al.* [13] proposed the first theoretical framework for measuring the vulnerability of biometric systems to attacks based on morphed biometric samples. Further, key factors which take a major influence on a system's vulnerability to such attacks have been identified, e.g. the shape of genuine and impostor score distributions or the FMR the system is operated at. To evaluate the vulnerability of biometric systems to attacks based on morphed images or templates, Scherhag *et al.* [14] introduced new metrics for vulnerability reporting, which strongly relate to the metrics defined in [15]. In addition, the authors provide recommendations on the assessment of morphing techniques. It is emphasized that unrealistic assumptions with respect to the quality of morphed biometric samples might cloud the picture regarding the performance of detection algorithms. It is important to note that so far there is no publicly available database of morphed face images and no publicly available morph detection algorithms.

In this work, the photo response non-uniformity (PRNU) is used to detect morphed face images. The PRNU [16] of an imaging sensor has emerged as an important tool for diverse forensic tasks including the detection of digital forgeries. It is shown that the proposed region-based analysis of PRNU behaviour reliably detects morphed face images. On a comprehensive database of bona fide and morphed face images practical detection performance is achieved. Moreover, we estimated the impact of different image post-processing steps applied to morphed face images on the detection performance of the proposed approach.

This paper is organized as follows: details on the employed extraction of PRNU signals are summarized in Sect. II. The proposed morph detection system is described in detail in Sect. III. Experimental results are presented in Sect. IV. Finally, conclusions are given in Sect. V.

II. PRNU extraction

The PRNU is a noise-like pattern, originating from slight variations among individual pixels during the conversion of photons to electrons in digital image sensors. It forms an inherent part of those sensors, whereas this weak signal is embedded into each and every image they capture.

This systemic and individual pattern is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering. The extraction of the PRNU noise residual from an image is performed by applying Fridrich's approach [17]. For each image I the noise residual W_I is estimated as described in Eq. (1),

$$W_I = I - F(I) \tag{1}$$

where F is a denoising function which filters out the sensor pattern noise. In this work, the denoising filter proposed by Mihcak *et al.* [18] is used in conjunction with a filtering distortion removal (FDR) PRNU enhancement proposed by Lin *et al.* [19]. Said enhancement aims at improving the SNR of the extracted PRNU noise residual W_I in a two step process by abandoning certain components that are severely contaminated by filtering errors introduced during the denoising of images. For further details on the denoising filter and FDR PRNU enhancement we refer to [18], [19]. Fig. 2 presents the extracted PRNU and FDR enhancement result for an exemplary image.

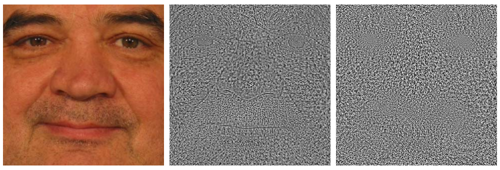
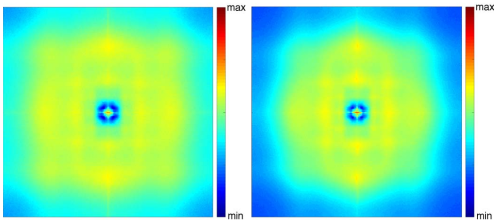

(a) Original (b) PRNU (c) FDR enh.

Fig. 2: Example of PRNU extraction and FDR enhancement for a pre-processed face image.



(a) Bona fide (b) Morphs

Fig. 3: DFT magnitude spectra of the PRNUs extracted from bona fide and morphed face images, averaged over the whole dataset.

The following essential criteria, which have been described by Fridrich *et al.* [20], make the PRNU well suited for the face morph detection scenario dealt with in this work:

1) *Universality*: all imaging sensors exhibit PRNU.
2) *Generality*: the PRNU is present in every picture independently of the scene content, with the exception of completely dark or overexposed images.
3) *Robustness*: it survives lossy compression, filtering, gamma correction, and many other typical processing procedures. It is even reported to survive high quality printing and scanning [21].

We decided to use the PRNU for the morphing detection, because it is unrelated to the image content and is present in every image acquired with a digital camera, as described above. Thus, it offers significant advantages over analysing other high-frequency image components.

By investigating the spectral characteristics of the PRNU it is possible to detect whether the images have been subject to further processing, e.g. non-geometrical operations have an influence on the strength of the PRNU [17]. By taking into consideration the processing steps applied during the face morphing, which consist of non-linear warping and averaging operations introducing interpolation artefacts, the distribution of the PRNU values is expected to change after such processing operations. Fig. 3 shows the discrete Fourier transform (DFT) magnitude spectra obtained by averaging the PRNU of all bona fide and morphed face images contained in the investigated dataset, which is described in Sect. IV. It clearly reveals a reduction of the high-frequency components within the DFT magnitude spectrum for the morphed images, as compared to the bona fide images. Furthermore, the spectrum is compressed, causing the area of the larger magnitudes to shrink. These effects are likely caused by the averaging and non-linear warping operations that occur during the morphing process and change the distribution of the DFT magnitudes.

Our approach aims at exploiting these effects in order to perform a blind no-reference face morph detection, which is presented in the following section.

III. Detection of morphed face images

As stated in the previous section, the goal of the proposed PRNU-based morph detection system is to exploit the spectral alterations introduced by the non-linear warping during the face morphing process within the PRNU to be able to discriminate between bona fide and morphed images. Furthermore, the discrimination is performed in a blind manner, i.e. without the need for any trusted bona fide reference image of one of the morphed subjects.

The proposed system follows the divide and conquer principle and consists of four major components: (A) *PRNU extraction*, (B) *PRNU splitting*, (C) *cell-wise feature extraction*, and (D) *cell aggregation*. The remainder of this section will discuss the different processing steps in more detail.

A. PRNU Extraction

The PRNU for each individual image is extracted, as described in Sect. II, by using the wavelet-based denoising filter by Mihcak *et al.* [18]. The extracted PRNU is then further enhanced using the FDR (frequency distortion removal) PRNU enhancement proposed by Lin *et al.* [19]. The PRNU is always extracted for the whole image, whereat every colour image is converted to grey-scale first according to [17]. The outcome of the PRNU extraction and PRNU enhancement process is illustrated in Fig. 2.

B. PRNU Splitting

The proposed system is able to work with the PRNU from the whole image, as well as arbitrary splits of the PRNU into multiple equisized cells. In this work, we investigate different cells configurations, from the whole image as a single cell up to $N = 10 \times 10$ cells. A larger number of cells is expected to further expose the non-linear transformations of the PRNU during the morphing process by putting stronger emphasis on local variations within an image. Eventually, we obtain N different cells $C_1, \ldots, C_N$. Fig. 4 shows an example of how the PRNU is split into $N = 2 \times 2$ equisized cells.

C. Cell-wise Feature Extraction

The feature extraction is performed for every cell individually. The first step consists in obtaining the frequency spectrum of the PRNU in each cell, which is done by means of the discrete Fourier transform (DFT). The resulting magnitude spectrum, as already shown in Sect. II, reveals the alterations of the PRNU signal caused by the morphing process. To quantify these effects, we calculate the histogram of the DFT magnitudes in order to represent the magnitude distribution within the spectrum. Fig. 5 shows the DTF magnitude spectra of a bona fide and morphed sample image with the corresponding histograms, where a shift of the magnitude distribution can be observed. All DFT magnitude histograms have been constrained to the same universal range of $[0, 8]$ and are divided into 100 bins. The range has been established with the values obtained from the DFT of all extracted PRNUs.

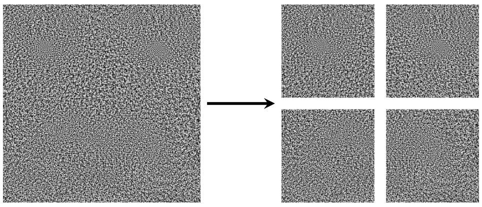


Fig. 4: Example for splitting the PRNU into $N = 4$ equisized cells (2×2).

Based on the observations from Sect. II, this magnitude histogram forms the basis for the different morph detection approaches in this work. We select the position of the peak P_{pos} in the histogram and its height or value P_{val} as being suited for the discrimination between bona fide and morphed images. We obtain P_{val} and P_{pos} as follows:

$$P_{val} = \max_{n=1\ldots b} H(n) \tag{2}$$

$$P_{pos} = \arg\max_{n=1\ldots b} H(n) \tag{3}$$

where b is the number of bins and H is the histogram of a cell. P_{pos} describes the position (bin) of the peak in the DFT magnitude histogram, while P_{val} represents the value (relative frequency) of the corresponding bin.

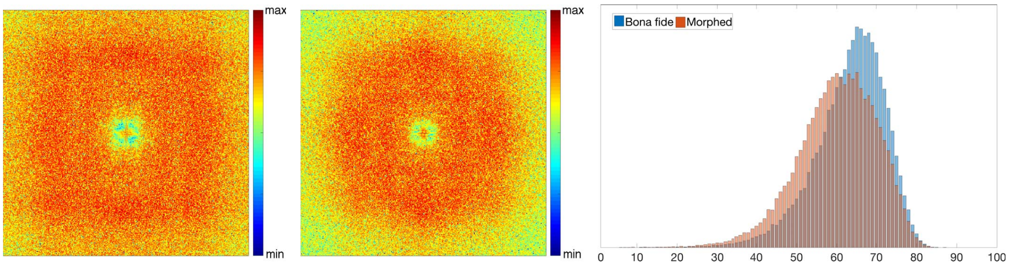
Furthermore, we consider the product of the peak position and value P_{pv} within the DFT magnitude histograms as a third combined feature:

$$P_{pv} = \max_{n=1\ldots b} H(n) * \arg\max_{n=1\ldots b} H(n) \tag{4}$$

Finally, we obtain a scalar value P for each PRNU cell, which is calculated using one of the the three approaches defined in Eqs. 2 to 4.

D. Cell Aggregation

As final step, the extracted features P for each cell C_n, in form of scalar values, are aggregated to obtain a global score S for the image. We investigated various strategies, whereas



(a) Bona fide image (b) Morphed image (c) DFT magnitude histograms

Fig. 5: Comparison of DFT magnitude spectra and histograms of a bona fide and a morphed sample image.

TABLE I: Database used for experimental evaluations

Gender	No. of subjects	No. of images	Bona fide images	Morphed images
Male	58	2,210	499	1,711
Female	39	1,165	462	703
All	97	3,375	961	2,414

we will present the two best performing ones. The aggregation strategies used in this work are:

$$S_{mean} = \frac{1}{N}\sum_{n=1}^{N} P_n \quad (5)$$

$$S_{rms} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} P_n^2} \quad (6)$$

where N is the number of total PRNU cells and P_n is the feature (scalar value) obtained for the PRNU cell C_n, as described in the previous processing step.

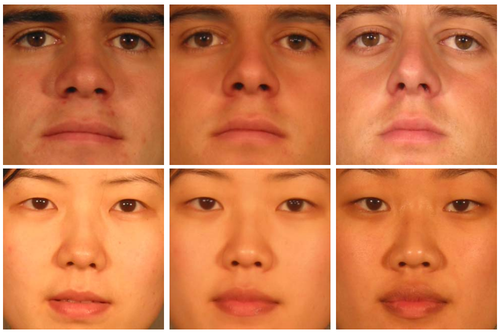
S_{mean} simply averages the scores of the individual cells, while S_{rms} characterizes the root mean square of the scores of all PRNU cells within an image. Eventually, we obtain a single scalar value S per image using one of the Eqs. 5 or 6. The value of S then indicates whether a face image has been created by morphing other face images or not. The final decision for a face image can be taken by a simple threshold.

IV. Experiments

In the following subsection, the generation of morphed face images and applied post-processing steps are described. In subsequent subsections, experimental results are reported which comprise a face recognition vulnerability assessment and a morph detection performance estimation.

A. Morph Generation and Post-processing

Experiments are performed on a subset of the FRGCv2 face database. A total number of 961 frontal faces with neutral expression have been manually selected and ICAO compliance has been verified, i.e. the distance between the eyes of a face has to be at least 90 pixels [22]. Details about the employed database are listed in Table I. In order to morph two face images the *dlib* facial landmark detector [23] is applied to both images. Subsequently, a Delaunay triangulation is performed to the average of corresponding points. An affine transform is then applied to the sets of triangles in both face images resulting in two warped images which are alpha blended using a alpha value of 0.5. In the pre-processing stage an image is segmented and normalized according to eye coordinates detected by the landmark detector. Subsequently, the normalized region is cropped to 320×320 pixels using predefined offsets to ensure that the morph detection algorithm is only applied to the facial region. Based on this subset 2,414 morphed faces have been automatically generated for pairs of subjects of same gender using the *OpenCV* library. Example images of bona fide and morphed face images are shown in Fig. 6, which illustrates the high quality of morphed face images being well in the quality limits set forth by ICAO and ISO/IEC standards.



(a) Subject 1 (b) Morph (c) Subject 2

Fig. 6: Examples of bona fide and morphed face images of subjects of same gender, ethnicity and age group

In addition, we also investigate the robustness of the proposed morphing detection system against different post-processing techniques. For this work we investigate four dif-

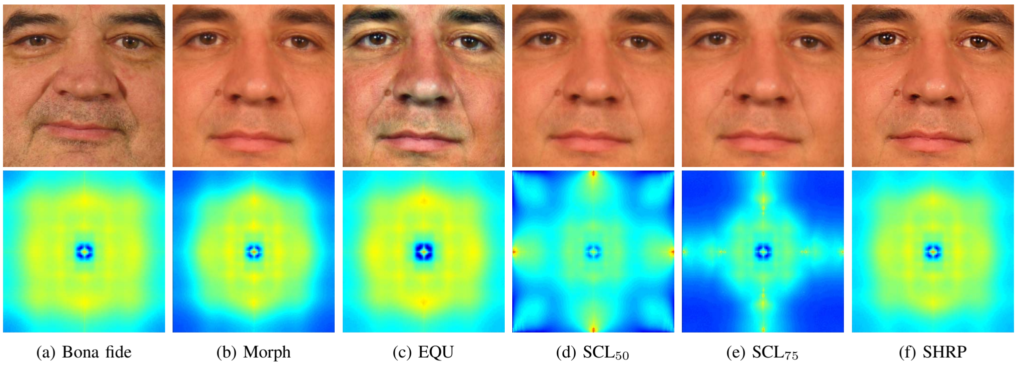


Fig. 7: Bonafide image (a) and results of applying the different post-processings to a morphed image (b to f). Below, the corresponding DFT magnitude spectra are shown (averaged over the whole dataset).

ferent techniques, which aim at further modifying the quality of the morphed face images:

- *EQU*: Contrast limited adaptive histogram equalization (CLAHE)
- *SCL*$_{50}$: Downscaling the image to 50% of its original size and subsequent upscaling
- *SCL*$_{75}$: Downscaling the image to 75% of its original size and subsequent upscaling
- *SHRP*: Sharpening the image using unsharp masking

The results of applying these post-processings to a morphed image and how they affect its DFT magnitude spectrum are demonstrated in Fig. 7.

B. Face Recognition Vulnerability Assessment

The attack success of the generated morphing attacks on a commercial-of-the-shelf face recognition system is evaluated using the metrics defined in [14]. In particular, the Relative Morph Match Rate (RMMR) and the ProdAvg Mated Morph Presentation Match Rate (ProdAvg-MMPMR).

When employing the default decision threshold of the COTS face recognition system a near-perfect MMPMR and RMMR (> 0.99) is obtained for using original morphed face images as well as post-processed. This means almost all face images of subjects contributing to a morphed face image are successfully matched against it which emphasizes the necessity of a robust morph detection subsystem. While the post-processings have a negligible impact on the vulnerability of the face recognition systems to morphing attacks, they should hamper the automatic detection of morphs.

C. Morph Detection Performance Evaluation

The performance of the detection algorithms is reported according to metrics defined in ISO/IEC 30107-3 [15]. The Attack Presentation Classification Error Rate (APCER) is defined as the proportion of attack presentations using the same presentation attack instrument species incorrectly classified as bona fide presentations in a specific scenario. The Bona Fide Presentation Classification Error Rate (BPCER) is defined as the proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario. The D-EER, i.e. the operation point where APCER $=$ BPCER, is used as general operation point and reported for images with and without post-processing. In addition, the BPCER10, i.e. the operation point where APCER $= 10\%$, and BPCER20, i.e. the operation point where APCER $= 5\%$, are estimated.

The performance of the proposed morph detectors is listed in Table II. The *Feature* column contains different combinations of extracted features P and aggregation strategies S, which are defined in Sect. III. The basic attempt using the whole image as a single cell, denoted as 1×1 in the table, is suitable to detect morphed face images with an D-EER as low as 2.1%. It is possible to improve the performance by splitting the image into cells, however, if the fragmentation is smaller than 8×8 cells, where a D-EER as low as 1.4% is achieved, the detection performance decreases again. Due to the lack of robustness to histogram shifts some post-processing techniques, e.g. equalization (*EQU*) and sharpening (*SHRP*), are severely influencing the performance of the algorithm. Note that depending on the direction of the histogram shift the results might even improve, as for *SCL*. This lack of robustness can be partially compensated for *SHRP* by employing a higher fragmentation of 8×8 cells, which is able to lower the D-EER to 11.9%. However, the *EQU* post-processing cannot be compensated at all. Clearly, further improvement of the detection algorithms is needed to counter this type of post-processing. The performance of the detectors highly depends on the type of aggregation (only the two best performing ones are presented in this work), as well as on the number of cells. On the given dataset the best overall performance was achieved with $P_{pos}|S_{mean}$ and $P_{pos}|S_{rms}$ with 8×8 cells (marked bold in Table II), yielding a D-EER as low as 2.2% on the original morphed images, 0.0% to 0.8% on scaled images

TABLE II: Performance of proposed PRNU-based morph detectors

		D-EER					BPCER10					BPCER20				
Feature	**Cells**	**Morph**	**EQU**	**SCL$_{50}$**	**SCL$_{75}$**	**SHRP**	**Morph**	**EQU**	**SCL$_{50}$**	**SCL$_{75}$**	**SHRP**	**Morph**	**EQU**	**SCL$_{50}$**	**SCL$_{75}$**	**SHRP**
$P_{val}\|S_{mean}$	1	2.1%	34.8%	0.7%	2.2%	46.4%	0.6%	52.2%	0.2%	0.7%	78.4%	1.1%	58.5%	0.3%	1.3%	85.9%
$P_{val}\|S_{rms}$		2.1%	34.8%	0.7%	2.2%	46.4%	0.6%	52.2%	0.2%	0.7%	78.4%	1.1%	58.5%	0.3%	1.3%	85.9%
$P_{pos}\|S_{mean}$		5.1%	36.4%	4.5%	0.3%	20.1%	1.5%	68.2%	1.4%	0.0%	39.4%	5.3%	77.4%	3.8%	0.0%	57.0%
$P_{pos}\|S_{rms}$		5.1%	36.4%	4.5%	0.3%	20.1%	1.5%	68.2%	1.4%	0.0%	39.4%	5.3%	77.4%	3.8%	0.0%	57.0%
$P_{pv}\|S_{mean}$		2.2%	32.9%	0.9%	0.2%	36.9%	0.2%	50.5%	0.1%	0.0%	64.3%	0.6%	59.1%	0.2%	0.0%	77.3%
$P_{pv}\|S_{rms}$		2.2%	32.9%	0.9%	0.2%	36.9%	0.2%	50.5%	0.1%	0.0%	64.3%	0.6%	59.1%	0.2%	0.0%	77.3%
$P_{val}\|S_{mean}$	2	2.0%	36.3%	0.7%	2.0%	45.8%	0.5%	53.2%	0.1%	0.7%	77.4%	1.0%	59.9%	0.3%	1.1%	84.3%
$P_{val}\|S_{rms}$		2.0%	36.3%	0.6%	2.0%	45.9%	0.5%	53.0%	0.1%	0.7%	77.7%	1.0%	59.8%	0.3%	1.1%	84.6%
$P_{pos}\|S_{mean}$		3.3%	33.4%	2.5%	0.2%	17.1%	0.9%	63.2%	0.8%	0.0%	31.3%	2.1%	74.3%	1.4%	0.0%	49.6%
$P_{pos}\|S_{rms}$		3.2%	33.1%	2.4%	0.2%	17.0%	0.8%	62.8%	0.7%	0.0%	31.1%	1.9%	73.5%	1.3%	0.0%	47.4%
$P_{pv}\|S_{mean}$		1.7%	32.8%	1.0%	0.1%	32.6%	0.4%	50.7%	0.1%	0.1%	60.7%	0.8%	60.3%	0.3%	0.1%	74.0%
$P_{pv}\|S_{rms}$		1.6%	32.6%	1.0%	0.1%	33.1%	0.4%	50.4%	0.1%	0.1%	61.0%	0.8%	60.0%	0.3%	0.1%	74.4%
$P_{val}\|S_{mean}$	4	1.9%	35.3%	0.5%	3.6%	40.5%	0.7%	51.4%	0.1%	1.7%	64.5%	1.1%	58.2%	0.2%	3.0%	72.8%
$P_{val}\|S_{rms}$		1.9%	35.1%	0.5%	3.6%	41.3%	0.7%	51.3%	0.1%	1.8%	66.0%	1.1%	58.3%	0.2%	2.9%	73.6%
$P_{pos}\|S_{mean}$		2.9%	33.0%	1.5%	0.1%	12.3%	0.2%	59.9%	0.1%	0.0%	16.5%	1.1%	71.8%	0.4%	0.0%	39.5%
$P_{pos}\|S_{rms}$		2.8%	32.8%	1.4%	0.1%	12.3%	0.2%	59.5%	0.1%	0.0%	16.6%	1.0%	71.4%	0.4%	0.0%	40.0%
$P_{pv}\|S_{mean}$		1.5%	32.3%	0.5%	0.1%	22.0%	0.2%	48.9%	0.0%	0.0%	41.0%	0.4%	57.8%	0.1%	0.0%	58.3%
$P_{pv}\|S_{rms}$		1.5%	32.0%	0.5%	0.1%	23.7%	0.2%	48.5%	0.0%	0.0%	43.7%	0.4%	57.6%	0.1%	0.0%	60.7%
$P_{val}\|S_{mean}$	8	3.2%	35.5%	0.4%	7.4%	34.5%	1.2%	53.5%	0.0%	6.2%	54.5%	2.1%	61.1%	0.1%	9.3%	64.1%
$P_{val}\|S_{rms}$		3.3%	35.6%	0.4%	7.6%	35.8%	1.3%	53.5%	0.0%	6.5%	56.7%	2.3%	61.0%	0.1%	10.1%	65.9%
$P_{pos}\|S_{mean}$		**2.2%**	**33.8%**	**0.7%**	**0.0%**	**10.8%**	**0.1%**	**60.2%**	**0.0%**	**0.0%**	**11.7%**	**0.6%**	**71.5%**	**0.1%**	**0.0%**	**30.8%**
$P_{pos}\|S_{rms}$		**2.3%**	**33.6%**	**0.8%**	**0.0%**	**11.0%**	**0.1%**	**59.8%**	**0.0%**	**0.0%**	**13.2%**	**0.6%**	**71.4%**	**0.1%**	**0.0%**	**32.8%**
$P_{pv}\|S_{mean}$		1.4%	31.8%	0.3%	0.1%	15.9%	0.2%	51.9%	0.0%	0.0%	24.0%	0.4%	60.5%	0.0%	0.0%	44.2%
$P_{pv}\|S_{rms}$		1.5%	31.3%	0.3%	0.0%	17.3%	0.2%	51.2%	0.0%	0.0%	26.9%	0.4%	60.2%	0.0%	0.0%	48.8%
$P_{val}\|S_{mean}$	10	3.8%	36.7%	0.3%	9.0%	33.1%	1.5%	54.5%	0.0%	8.3%	51.2%	3.0%	60.4%	0.1%	13.2%	61.3%
$P_{val}\|S_{rms}$		3.9%	36.7%	0.3%	9.3%	34.1%	1.6%	54.8%	0.1%	8.6%	53.2%	3.3%	60.7%	0.1%	14.3%	63.0%
$P_{pos}\|S_{mean}$		2.4%	34.9%	0.6%	0.0%	10.5%	0.0%	61.7%	0.0%	0.0%	11.2%	0.7%	71.6%	0.0%	0.0%	28.4%
$P_{pos}\|S_{rms}$		2.6%	34.6%	0.6%	0.0%	10.9%	0.0%	61.3%	0.0%	0.0%	12.3%	0.8%	71.6%	0.0%	0.0%	30.3%
$P_{pv}\|S_{mean}$		1.8%	32.8%	0.2%	0.1%	13.9%	0.1%	53.3%	0.0%	0.0%	20.8%	0.3%	62.7%	0.0%	0.0%	41.7%
$P_{pv}\|S_{rms}$		1.8%	32.4%	0.2%	0.0%	15.0%	0.1%	52.7%	0.0%	0.0%	25.0%	0.3%	62.3%	0.0%	0.0%	46.1%

and as low as 10.8% on sharpened images. An appropriate choice for the amount of used cells obviously relates on the resolution of the processed image. Overall, it can be observed that both aggregation strategies S_{mean} and S_{rms} obtain similar results across all extracted features. Furthermore, a higher fragmentation of up to 8 × 8 cells, and therefore analysis of the local alterations within the image, is observed to be beneficial to the detection performance. The position of the peak P_{pos} in the DFT magnitude spectrum emerged as the most stable among the extracted features across all applied post-processings.

V. Conclusion and Future Work

In this work, we proposed an automated morph detection for face images based on the PRNU. The procedure of creating morphed face images takes influence on the property of PRNU values, in particular across different image regions. It is shown that a cell-based PRNU analysis allows for a reliable detection of morphed face images. Furthermore, we analysed the impact of different image post-processing techniques on the detection performance, where the proposed detection system was robust against scaling and sharpening of the images, and only failed for the applied histogram equalisation. Deeper investigation and an improvement of the detection approaches is clearly needed to counter the failed detection of morphed images in this case.

Future studies might also include a vulnerability analysis of proposed detection algorithms to attacks based on PRNU insertion/substitution. Additionally, an investigation of the proposed morph detection systems for images from different cameras as well as for printed and scanned images could be subject to future work.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the Center for Research in Security and Privacy (CRISP). Furthermore, this work was also supported by the Austrian Science Fund (FWF) under Project No. P26630.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] S. Z. Li and A. K. Jain, *Handbook of Face Recognition (2nd edition)*. Springer, 2011.

[3] "FRONTEX – Research and Development Unit: Best practice technical guidelines for automated border control (ABC) systems," 2012, version 2.0.

[4] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[5] S. Marcel, M. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing*. Springer-Verlag New York, Inc., 2014.

[6] ISO/IEC TC JTC1 SC37 Biometrics, *ISO/IEC IS 30107-1. Information Technology – Biometrics presentation attack detection – Part 1: Framework*, International Organization for Standardization, Mar. 2016.

[7] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. Int. Joint Conf. on Biometrics (IJCB)*, 2014, pp. 1–7.

[8] U. Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, “On the vulnerability of face recognition systems towards morphed face attacks,” in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.

[9] M. Ferrara, A. Franco, and D. Maltoni, “On the effects of image alterations on face recognition accuracy,” in *Face Recognition Across the Imaging Spectrum*, T. Bourlai, Ed. Springer International Publishing, 2016, pp. 195–222.

[10] M. Hildebrandt, T. Neubert, A. Makrushin, and J. Dittmann, “Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps,” in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.

[11] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann, “Modeling attacks on photo-ID documents and applying media forensics for the detection of facial morphing,” in *Proc. Workshop on Information Hiding and Multimedia Security (IH& MMSec)*, 2017, pp. 21–32.

[12] R. Ramachandra, K. Raja, S. Venkatesh, and C. Busch, “Transferable deep-cnn features for detecting digital and print-scanned morphed face images,” in *2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW)*, July 2017.

[13] M. Gomez-Barrero, C. Rathgeb, U. Scherhag, and C. Busch, “Is your biometric system robust to morphing attacks?” in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.

[14] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Raghavendra, and C. Busch, “Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting,” in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2017, pp. 1–12.

[15] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC IS 30107-3:2017, IT – Biometric presentation attack detection – Part 3: Testing and Reporting*.

[16] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, “Determining image origin and integrity using sensor noise,” *Trans. Info. For. Sec.*, vol. 3, no. 1, pp. 74–90, Mar. 2008.

[17] J. Fridrich, “Digital image forensic using sensor noise,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, March 2009.

[18] M. Mihcak, I. Kozintsev, and K. Ramchandran, “Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising,” in *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP ’99*. Phoenix, AZ, USA: IEEE, Mar. 2009, pp. 3253–3256.

[19] X. Lin and C.-T. Li, “Enhancing sensor pattern noise via filtering distortion removal,” *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 381–385, 2016.

[20] J. Fridrich, “Sensor defects in digital image forensics,” in *Digital Image Forensics: There is more to a picture than meets the eye*, H. Sencar and N. Memon, Eds. Springer Verlag, 2012, ch. 6, pp. 179–218.

[21] M. Goljan, J. Fridrich, and J. Lukas, “Camera identification from printed images,” in *Proceedings of SPIE, Electronic Imaging, Forensics, Security, Steganography, and Watermarking of Multimedia Contents X*. San Jose, CA, USA: SPIE, Jan. 2008.

[22] International Organization for Standardization, “Information technology – Biometric data interchange formats – Part 5: Face image data,” JTC 1/SC 37, ISO/IEC 19794-5:2005 consolidated, 2005.

[23] D. E. King, “Dlib-ml: A machine learning toolkit,” *Journal of Machine Learning Research*, vol. 10, 2009.

PRNU Variance Analysis for Morphed Face Image Detection

Luca Debiasi[*], Christian Rathgeb[†], Ulrich Scherhag[†], Andreas Uhl[*] and Christoph Busch[†]

[*]WaveLab – The Multimedia Signal Processing and Security Lab, Universität Salzburg, Austria
[†]da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
{ldebasi,uhl}@cs.sbg.ac.at,
{ulrich.scherhag,christian.rathgeb,christoph.busch}@h-da.de

Abstract

In this work, a method to detect morphed face images based on Photo Response Non-Uniformity (PRNU) is presented. More specifically, the variance of PRNU-based features across image cells is estimated to distinguish bona fide from morphed and potentially post-processed morphed face images. The proposed morph detector is shown to be robust against post-processing techniques, which are likely to be applied to conceal the morphing process, e.g. histogram equalisation or image sharpening. Tested on a database of 961 bona fide and 2,414 automatically morphed face images, a detection equal error rate (D-EER) of 10.5% is obtained over all investigated attacks, including unaltered morphed images and various post-processing techniques.

1. Introduction

Automated face recognition [36, 17] represents a long-standing field of research in which a major break-through has been achieved by the introduction of deep neural networks [33, 24]. Resulting performance improvements paved the way for deployments of face recognition technologies in diverse application scenarios, ranging from mobile device access control to Automated Border Control (ABC). However, recently researchers found that the intended generalisability of deep face recognition systems also increases their vulnerability against attacks, e.g. spoofing attacks (a.k.a. presentation attacks) [22]. Most notably, a specific attack against face recognition systems based on morphed face images has been proposed in [3].

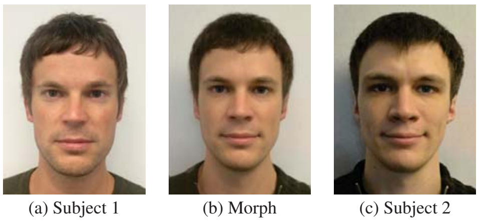

(a) Subject 1 (b) Morph (c) Subject 2

Figure 1: Examples for bona fide and morphed face images

Morphing techniques can be used to create artificial biometric samples, which resemble the biometric information of two (or more) individuals in image and feature domain. In order to morph two face images, an attacker usually defines corresponding landmarks and a triangulation of landmarks is done on both images. The landmarks are then averaged to a single set of landmarks and both images are warped according to the resulting triangulation. Finally, alpha-blending is performed. Realistic morphed face images can be generated by non-experts employing easy-to-use face morphing software which can be purchased at a reasonable price, e.g. FantaMorph[1]. Fig. 1 depicts an example of morphing two face images.

It has been shown that morphed face images are realistic enough to fool human examiners [4]. This means, there is a risk that morphed biometric images are infiltrated to a biometric recognition system at enrolment, e.g. during the issuance process of electronic travel documents. In [3] commercial face recognition software tools have been exposed to be highly vulnerable to attacks based on morphed face images. This means that the subjects contributing to the morphed image were both (or all) successfully matched against that single enrolled morphed image. These findings have been confirmed by other researchers, e.g. in [32]. In their vulnerability analysis, researchers used decision thresholds yielding a False Match Rate (FMR) of 0.1%, following the guidelines provided by the European Agency for the Management of Operational Cooperation at the Ex-

[1] FantaMorph: http://www.fantamorph.com/

ternal Borders (FRONTEX) [1].

In the recent past, researchers have presented different approaches to distinguish bona fide from morphed face images, see Sect. 2. Proposed approaches either processes a single potentially morphed image, i.e. *no-reference* morph detection, or a potential morph together with a trusted live capture from an authentication attempt, i.e. *differential* morph detection. In the no-reference scenario different media forensic concepts have been applied [23, 16, 2]. Adaptations of such techniques, which are designed to detect digital forgeries, revealed promising results for the detection morphed face images. In particular, a PRNU-based detection of morphed face images was introduced in [2]. The extraction of the PRNU and an analysis of its distributions across image cells has been reported to reliably detect morphed face images, while the approach fails if image post-processing, e.g. histogram equalisation, is applied to generated morphs.

The work presented in this paper was inspired by the approach of [2] and proposes a PRNU variance analysis for morphed face image detection. It is shown that an increased variance of different PRNU statistics across image cells is a reliable indicator for image morphing. Further, the improved PRNU-based morph detector is shown to be resistant against common image post-processing methods. Finally, the presented approach is expected to be more robust against arbitrary post-processings, since it analyses image block interrelations rather than image features which might specifically result from a distinct morphing process applied to a certain face database.

This paper is organized as follows: related works are briefly discussed in Sect. 2. Fundamentals of PRNU extraction are explained in Sect. 3. The proposed morph detection method is described in detail in Sect. 4. Experimental results are reported in Sect. 5. Finally, conclusions and future works are summarized in Sect. 6.

2. Related Work

The topic of face morph detection has sparked the interest of numerous research laboratories working in the field of biometrics. Efforts to define evaluation metrics for morph detection and vulnerability analysis have already been made [28, 10], see Sect. 5. A recent overview on conducted vulnerability analyses and morph detection methods can be found in [20]. Presented approaches can be coarsely categorized with respect to the considered morph detection scenario. The majority of works assume the challenging no-reference scenario while some implement a differential morph detection which is motivated by the fact that trusted live captures are available in ABC scenarios.

A differential morph detection method referred to as demorphing was proposed in [5]. Within this approach a trusted live capture is aligned to a potential morph and subtracted from it in the image domain. The resulting image is then compared against the trusted live capture. A morph is detected if the biometric decision changes from "accept" to "reject". Robustness of de-morphing against slight face pose variations has been confirmed in [6]. Nevertheless, the authors note that in an ABC scenario the performance of de-morphing might degrade due to potential variations of quality and environmental conditions.

Several researchers have suggested the use of general purpose texture descriptors, e.g. Local Binary Patterns (LPB) or Binarized Statistical Image Features (BSIF), which have been employed widely for biometric recognition. Machine learning-based classifiers, e.g. Support Vector Machines (SVMs), are either trained directly on extracted feature vectors for no-reference morph detection [25, 29, 14] or differences between feature vectors can (additionally) be employed in a differential scenario [32]. Also, face-specific features such as differences between landmark positions or angles could be employed in a differential scenario which so far has been shown to reveal rather moderate detection performance [27]. Depending on the feature representation of texture descriptors the inputs of classifiers have to be adapted, e.g. for Scale-Invariant Feature Transform (SIFT) the number of extracted keypoints has been shown to be suitable for the task of morph detection [16, 32]. Score level fusions of different types of features have been proposed, too [30]. In particular, in the no-reference scenario classifiers may overfit to distinct micro texture features. These can be dataset-specific features which are altered or introduced by the applied morphing process. It has been shown that the performance of morph detectors based on general purpose texture descriptors might significantly decrease if training and test images stem from a different source, i.e. face database [31].

The use of convolutional neuronal networks for no-reference morphed face detection has been proposed by different researchers reporting promising results [26, 35]. Again, with these methods there is potentially a problem of overfitting. In particular, resulting deep classifiers may favour image locations where artefacts, e.g. shadows around the iris region, are likely to appear due to an imperfect automated morph creation process. Further, published approaches have been trained and tested for a single morph generation method, i.e. generalizability still has to be evaluated.

Focusing on the no-reference scenario diverse approaches related to media forensics have been presented. In different works, the detection of JPEG double-compression artefacts has been suggested for the purpose of morph detection [19, 10, 20]. However, the presence of such artefacts implies a strong assumption on the image format of face images used for morph generation as well as the resulting morphed face image. The International Civil Aviation Organization (ICAO) suggests face image data to be stored in accordance with the specifications established by the International Organization for Standardization (ISO) in [12]. More specifically, the ICAO recommends face images to be stored in electronic travel documents at an average compressed sizes of 15kB to 20kB in JPEG or JPEG 2000 format [11]. Hence, depending on the image size and the employed compression algorithm the detection of JPEG double-compression artefacts might not be feasible. In [34] a morph detection method based on reflection analysis in face images is presented. The lightning direction is estimated based on reflections detected in the eyes of a potentially morphed image. Subsequently, reflections on the nose of the face are analysed. However, ISO requires hot spots and specular reflections to be absent in face images used in electronic travel documents. In particular, diffused lighting, multiple balanced sources or other lighting methods shall be used, i.e. a single bare "point" light source like a camera mounted flash is not acceptable for imaging [12]. Morph detection methods based on continuous image degradation have been proposed in [23, 16]. The basic idea behind these methods is to continuously degrade the image quality, e.g. by using JPEG compression, to create multiple artificial self-references of a face image. The distances from these references to the original image are then analysed for morph detection. Additionally, PRNU-based morph detection has been proposed in [2]. This approach is described in more detail in Sect. 4.

Despite promising results reported in many works a reliable detection of morphed face images still represents an open research challenge. Note that the generalizability/robustness of published approaches has not been shown, as these have been mostly trained and tested on single databases using a single morph generation algorithm. Further, the likely application of image post-processing techniques, e.g. image sharpening, is neglected in most works. Lastly, so far there are no publicly available database of bona fide and morphed face images and no publicly available morph detection algorithms.

3. PRNU Extraction and Characteristics

Digital image forensics aims at acquiring knowledge on visual contents and acquisition devices by evaluating the traces that are left on the data during the acquisition and in the subsequent processing. The PRNU of imaging sensors [7] emerged as an important forensic tool. It can be used for a variety of important tasks, such as device identification, device linking, recovery of processing history, and detection of digital forgeries. The PRNU is an intrinsic property of all digital imaging sensors, which is characterised by slight variations among individual pixels in their ability to convert photons to electrons. Consequently, every sensor casts a weak noise-like pattern onto every image it captures. This noise-like pattern can be considered as an unintentional stochastic spread-spectrum watermark.

In [7] Fridrich presents an approach on how to extract the PRNU noise residual from an image. For each image I the noise residual W_I is estimated as described in Eq. (1),

$$W_I = I - F(I) \qquad (1)$$

where F is a denoising function which filters out the sensor pattern noise. In this work, the denoising filter proposed by Mihcak *et al*. [21] is used in conjunction with a Filtering Distortion Removal (FDR) PRNU enhancement proposed by Lin *et al*. [18]. Said enhancement aims at improving the SNR of the extracted PRNU noise residual W_I in a two step process by abandoning certain components that are severely contamined by filtering errors introduced during the denoising of images. For further details on the denoising filter and FDR PRNU enhancement we refer to [21, 18]. Fig. 2 shows the extracted and enhanced PRNU for an exemplary face image.

The PRNU offers some essential advantages for the detection of morphed face images. First of all, as stated by Fridrich *et al*. [8], all digital image sensors exhibit PRNU, which makes this sensor noise virtually present in every captured image. Furthermore, it is independent from the scene content and even robust against typical processing procedures like lossy compression or gamma correction, and it is even reported to be robust against high quality printing and scanning [9].

These criteria make the PRNU well suited for the morph detection scenario investigated in this work, because it offers significant advantages over analysing other high-frequency image components: First and foremost the PRNU is present in every image acquired with a digital camera, hence virtually every face image. In addition, in principle the PRNU is unrelated to the image content, but its high-frequency components might interfere with the PRNU. However, this interference can be attenuated by different PRNU enhancement approaches.

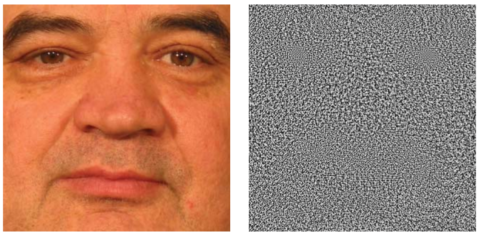

(a) Original (b) PRNU

Figure 2: Extracted and enhanced PRNU for an exemplary face image.

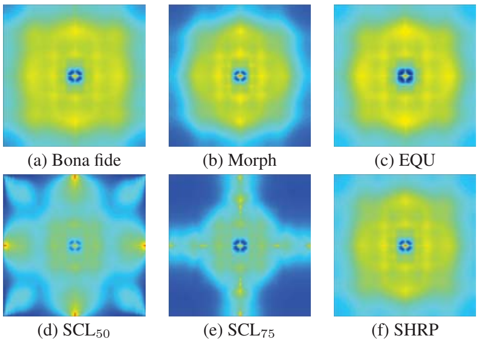


Figure 3: Averaged PRNU DFT magnitude spectra of bona fide images (a), morphed images (b) and post-processed morphed images (c to f).

The spectral characteristics of the PRNU reveal whether an image has been subject to further processing [7]. Since face morphing usually comprises different non-linear warping and averaging operations, the distribution of the PRNU values is affected by these operations, as previously shown in [2]. The PRNU's DFT magnitude spectrum of morphed images shows a reduction of the high-frequency components as well as a compression of the whole spectrum, which is illustrated in Fig. 3b.

Debiasi *et al*. [2] furthermore investigated the effects of various post-processings on the PRNU's DFT magnitude spectrum. They applied four different post-processings to the morphed face images: Histogram equalisation (*EQU*), downscaling and subsequent upscaling (*SCL*$_{50}$, *SCL*$_{75}$) and sharpening (*SHRP*). More details are given in Sect. 5, while the effects of these operations are presented in Fig. 3. One can observe that the DFT spectra of *SCL*$_{50}$ and *SCL*$_{75}$ are clearly discriminable from bona fide images, whereas the spectra of *SHRP* and especially *EQU* show a high similarity to bona fide images.

4. Detection of Morphed Face Images

The PRNU-based morph detection system proposed by Debiasi *et al*. in [2] aims at exploiting the spectral alterations of the PRNU introduced by the non-linear warping during the face morphing process and therefore discriminate between bona fide and morphed images. Furthermore, the discrimination is performed in no-reference manner.

The morph detection system consists of five major components: (A) *PRNU extraction*, (B) *PRNU splitting*, (C) *cell-wise feature extraction*, (D) *cell aggregation* and the (E) *decision*. In short, the PRNU is extracted from a face image and divided into cells. Thereafter, the DFT magnitude spectrum is computed for each cell, whereof different features P are derived. By averaging the extracted features for each cell an aggregated score S is obtained. Finally, the system performs a binary decision (bona fide or morphed) based on a simple threshold, which can be determined by analysing the score distribution of bona fide images.

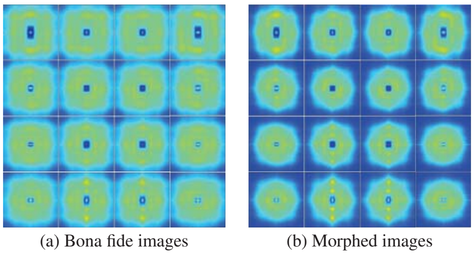


Figure 4: Illustration of variations across DFT magnitude spectra in morphed images compared to bona fide ones for 4×4 image cells (average of all images in dataset).

4.1. Variance Analysis

In this work, the approach of [2] is extended by proposing an analysis of the PRNU variance for morphed face image detection. Due to the morphing process's nature of producing inhomogeneous alterations across different image regions, an increased variance of the PRNU signal is expected across image cells. Fig. 4 shows the variations of the DFT magnitudes across different image cells of bona fide and morphed images. These local variations can be useful as a reliable indicator for image morphing. In order to analyse the variance of the PRNU, we propose some adaptations to Debiasi *et al*.'s [2] approach, which are presented in the remainder of this section. The proposed system is illustrated in Fig. 5.

4.1.1 Feature Extraction

In this work we propose to analyse the variance of two distinct features: P_{pos} and P_{en}. The first one, P_{pos}, has been proposed in [2] and is based on the PRNU's DFT magnitude histogram. It represents the peak's position (bin) within the histogram and is obtained as follows:

$$P_{pos} = \arg\max_{n=1\dots b} H(n), \quad P_{en} = \sum_{x \in M} |x|^2 \qquad (2)$$

where b is the number of bins and H is the magnitude histogram of a cell. As the second feature, P_{en}, we propose to compute the energy of the PRNU's DFT magnitudes, as defined in Eq. 2, where M are the DFT magnitudes within a cell and x their respective values. Both features lead to a scalar value P for each PRNU cell.

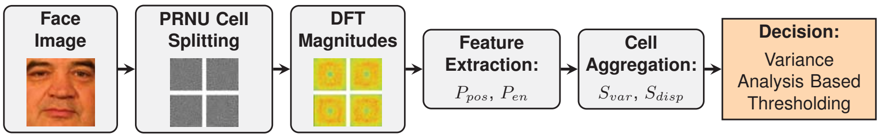

Face Image
PRNU Cell Splitting
DFT Magnitudes
Feature Extraction: P_{pos}, P_{en}
Cell Aggregation: S_{var}, S_{disp}
Decision: Variance Analysis Based Thresholding

Figure 5: Processing steps of the proposed PRNU-based morph detection system.

4.1.2 Cell Aggregation

In order to perform the variance analysis across all image cells, we make use of two measures of dispersion. The variance, S_{var}, is given by

$$S_{var} = Var(P) = \frac{1}{N}\sum_{n=1}^{N}(P_n - \bar{P})^2 \qquad (3)$$

$$\bar{P} = \frac{1}{N}\sum_{n=1}^{N} P_n, \quad S_{disp} = \frac{Var(P)}{\bar{P}} \qquad (4)$$

The index of dispersion, S_{disp}, or variance to mean ratio, is given in Eq. 4, where N is the number of total PRNU cells, P_n is the feature (scalar value) obtained for the PRNU cell C_n, as described previously, and $\bar{P}$ is the average feature value for all PRNU cells C. In both cases, we obtain a single scalar value S for each image.

4.2. Decision

As mentioned above, the PRNU-based morph detection system proposed in [2] makes use of a simple thresholding to determine if the presented image is a bona fide one or not. It was shown that with this one dimensional decision it was not possible to reliably detect some of the post-processed morphed images, i.e. *SHRP* and in particular *EQU*.

Due to the large variety of possible unknown post-processings, we decided to focus on the known properties of bona fide images and to use this knowledge to our advantage by simply deriving the mean variation $\bar{B}$ from the bona fide images. With this characteristic of bona fide images, we are able to calculate the distance D of an investigated image to bona fide images as

$$D = |S - \bar{B}|, \quad \bar{B} = \frac{1}{N_B}\sum_{n=1}^{N_B} S \qquad (5)$$

where S is the result of the cell-aggregation, $\bar{B}$ is the mean variation of all bona fide images N_B. The variation is either S_{var} or S_{disp}, whichever is used in the cell aggregation processing step. The final decision for a presented face image is taken by thresholding the distance D.

5. Experiments

In the following section, we describe the morphed face data set investigated in this work. In addition, we report experimental results which comprise a morph detection performance estimation and robustness under the presence of common post-processing techniques.

5.1. Face Morphing Data Set

In order to allow a direct comparison of the morph detection performance with [2], experiments are performed on a subset of the FRGCv2 face database, where 961 frontal faces with neutral expression have been manually selected as bona fide samples, which are all ICAO compliant according to [12]. Two face images are morphed by applying the *dlib* facial landmark detector [15] to both images. Subsequently, a Delaunay triangulation is computed, which forms the basis for a subsequent affine transform to the sets of triangles in both face images. The final morphed image is generated by alpha blending of the two warped images using an alpha value of 0.5.

The face images are then segmented and normalized according to eye coordinates detected by the *dlib* landmark detector. The resulting normalised region of interest is cropped to 320×320 pixels, to ascertain that the morphing detection algorithm is only applied to the facial region.

In total, $2,414$ high quality morphed face images have

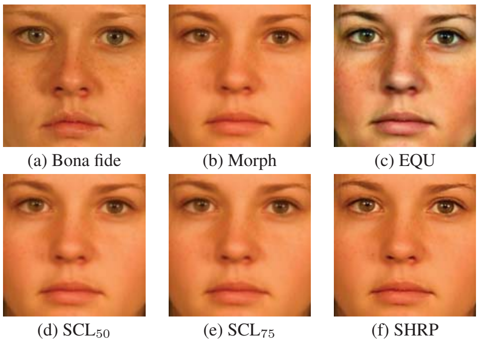

(a) Bona fide
(b) Morph
(c) EQU
(d) SCL$_{50}$
(e) SCL$_{75}$
(f) SHRP

Figure 6: Data set examples: bona fide image (a), morphed image (b) and post-processed morphs (c - f).

Table 1: Performance of proposed PRNU-based morph detectors

		D-EER					**BPCER10**					**BPCER20**				
Algorithm	**Cells**	**Morph**	**EQU**	**SCL$_{50}$**	**SCL$_{75}$**	**SHRP**	**Morph**	**EQU**	**SCL$_{50}$**	**SCL$_{75}$**	**SHRP**	**Morph**	**EQU**	**SCL$_{50}$**	**SCL$_{75}$**	**SHRP**
$P_{pos}\|S_{mean}$	4	2.9%	33.0%	1.5%	0.1%	12.2%	0.2%	59.9%	0.1%	0.0%	16.5%	1.1%	71.8%	0.4%	0.0%	39.5%
$P_{pos}\|S_{var}$		30.3%	49.6%	18.2%	42.0%	14.0%	75.1%	88.2%	45.5%	85.0%	24.9%	87.1%	94.0%	71.1%	91.6%	54.4%
$P_{pos}\|S_{disp}$		25.2%	49.7%	14.1%	33.3%	12.9%	64.1%	88.6%	27.6%	77.3%	21.4%	79.9%	93.9%	57.5%	87.3%	50.9%
$P_{en}\|S_{var}$		19.4%	29.5%	4.3%	9.0%	2.3%	47.8%	51.6%	1.3%	7.9%	0.1%	69.3%	64.6%	3.5%	19.0%	0.6%
$P_{en}\|S_{disp}$		15.3%	30.3%	3.4%	5.5%	2.5%	30.2%	53.5%	0.6%	2.9%	0.1%	54.6%	67.0%	2.1%	6.1%	0.8%
$P_{pos}\|S_{mean}$	8	2.2%	33.8%	0.7%	0.0%	10.8%	0.1%	60.2%	0.0%	0.0%	11.7%	0.6%	71.5%	0.1%	0.0%	30.8%
$P_{pos}\|S_{var}$		18.5%	49.8%	2.5%	34.9%	4.9%	36.7%	89.4%	0.7%	78.6%	1.3%	64.6%	94.4%	1.3%	88.3%	4.7%
$P_{pos}\|S_{disp}$		11.1%	49.8%	1.7%	16.5%	4.8%	12.4%	89.6%	0.1%	29.8%	1.2%	27.7%	94.5%	0.7%	51.3%	4.1%
$P_{en}\|S_{var}$		20.2%	15.8%	4.2%	11.0%	1.3%	44.6%	20.3%	1.5%	12.3%	0.0%	66.0%	30.1%	3.5%	24.7%	0.2%
$P_{en}\|S_{disp}$		12.7%	16.8%	2.9%	4.5%	1.6%	16.2%	23.0%	0.5%	2.4%	0.0%	33.9%	33.1%	1.6%	4.1%	0.4%
$P_{pos}\|S_{mean}$	10	2.4%	34.9%	0.6%	0.0%	10.5%	0.0%	61.7%	0.0%	0.0%	11.2%	0.7%	71.6%	0.0%	0.0%	28.4%
$P_{pos}\|S_{var}$		15.3%	50.0%	1.4%	32.2%	3.6%	25.9%	90.0%	0.1%	77.6%	0.9%	44.2%	95.0%	0.4%	88.7%	2.2%
$P_{pos}\|S_{disp}$		7.5%	50.0%	1.0%	11.9%	3.8%	5.4%	90.0%	0.1%	15.1%	1.0%	11.8%	95.0%	0.1%	27.8%	2.2%
$P_{en}\|S_{var}$		18.3%	14.5%	3.5%	9.2%	1.1%	36.5%	17.5%	0.6%	8.3%	0.0%	56.4%	24.5%	2.3%	17.7%	0.0%
$P_{en}\|S_{disp}$		**11.0%**	**15.9%**	**2.6%**	**3.8%**	**1.5%**	**11.9%**	**20.0%**	**0.1%**	**1.9%**	**0.0%**	**22.0%**	**29.0%**	**0.9%**	**3.1%**	**0.1%**

been automatically generated for pairs of subjects of same gender using the *OpenCV* library, which are well within the quality limits defined by ICAO and ISO/IEC standards. Furthermore, Debiasi *et al*. [2] reported that the morphed face images generated for this data set pose a severe risk for a COTS face recognition system, since probe face images from both contributing subjects can match with the morph at high success rate. They obtained a Relative Morph Match Rate (RMMR) and the ProdAvg Mated Morph Presentation Match Rate (ProdAvg-MMPMR) of > 0.99, which emphasises the necessity of a robust morph detection system. For more details on metrics for reporting the vulnerability of face recognition systems to morphed faces, the reader is referred to [28].

Moreover, the data set also includes a variety of different post-processing techniques applied to the morphed images: *EQU*, *SCL*$_{50}$, *SCL*$_{75}$ and *SHRP*. They aim at hampering the detection performance of the morph detection system. Some examples for post-processed morphs, which are part of the investigated data set, are shown in Fig 6.

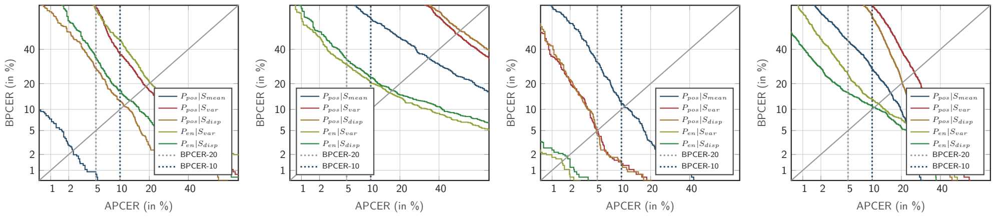
5.2. Morph Detection Performance Evaluation

The morph detection performance is examined according to metrics defined in ISO/IEC 30107-3 [13]: Attack Presentation Classification Error Rate (APCER) and bona fide Presentation Classification Error Rate (BPCER). APCER reports the proportion of attack presentations incorrectly classified as bona fide presentations in a specific scenario. BPCER, on the other hand, reports the proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario. The operation point of the system, where APCER = BPCER, is defined as detection equal error rate D-EER. Furthermore, two additional operation points, BPCER10 (where APCER $= 10\%$) and BPCER20 (where APCER $= 5\%$), are reported.

Tab. 1 summarises the obtained morph detection performance in form of D-EER, BPCER10 and BPCER20 for images with (*EQU*, *SCL*$_{50}$, *SCL*$_{75}$, *SHRP*) and without post-processing (*Morph*). The column *Algorithm* comprises the combinations of extracted features P and aggregation strategies S defined in Sect. 4. The column *Cells* contains the cell splits of the investigated images. We focused on cell splits of 4×4, 8×8 and 10×10 in this work due to the improved results with higher cell counts reported in [2].

The proposed algorithm by Debiasi *et al.* in [2], $P_{pos}|S_{mean}$ for 8×8 cells, serves as baseline and achieves a D-EER performance of 2.2% for unaltered morphs, but fails at detecting morphs post-processed with *EQU* at 33.8% and shows a high performance decrease for detecting sharpened morphs (*SHRP*) at 10.5%. Because the magnitude spectra of *SCL*$_{50}$, *SCL*$_{75}$ and *SHRP* post-processing are quite distinct bona fide image's ones, as it can be observed in Fig. 3, they can be detected quite reliably in general. The remaining algorithms are based on the variance analysis described in Sect. 4.1.

The proposed $P_{pos}|S_{var}$ and $P_{pos}|S_{disp}$ algorithms show rather inconsistent results among the different post-processings, especially they completely fail at detecting *EQU* morphs. Since they are based on the DFT magnitude histograms, they are highly vulnerable to histogram shifts such as those caused by histogram equalisation (*EQU*), leading to a D-EER of up to 50%. When looking at $P_{en}|S_{var}$ and $P_{en}|S_{disp}$, one can immediately note the degradation in unaltered morph detection of 11% in the best case (compared to the baseline of 2.2%), as shown in Fig. 7a. However, a more stable performance across all post-processed morphs is achieved. The highest performance gains are achieved for *EQU* and *SHRP* with a D-EER of 14.5% and 1.1% respectively, as compared to the baseline of 33.0% and 10.5%, which are illustrated in Fig. 7b and 7c. In general, the variance analysis based algorithms lead to a trade off between unaltered morph detection and post-



(a) Morph (b) EQU (c) SHRP (d) ALL

Figure 7: DET curves for PRNU-based morph detectors (10 × 10 cells).

processed morph detection. It enables the system to be more robust against different attacks, while also increasing the overall performance when all attacks are considered (*Morphs*, *EQU*, *SCL*$_{50}$, *SCL*$_{75}$, *SHRP*). This can mainly be attributed to the statistical variations caused by the morphing procedure across the image, which are most prominent in the P_{en} feature and can be captured best by the S_{disp} cell-aggregation strategy.

Thus, the overall best performing and most stable algorithm is based on the proposed variance analysis, $P_{en}|S_{disp}$ for 10 × 10 cells, is able to achieve respectable results across all altered and unaltered morphed images and is robust against a wide variety of post-processing attacks aiming at deteriorating the morph detection system. This robustness is a significant improvement over the baseline algorithm proposed in [2], which is much more vulnerable to post-processing attacks. Furthermore, the overall system performance is also improved from 15.7% average D-EER (baseline) to 10.5% D-EER (proposed algorithm), when all altered and unaltered morphs are considered. A direct comparison of both algorithms is presented in Fig. 7d and Tab. 2, where it can be observed that both algorithms have opposing strengths and weaknesses regarding the single post-processing techniques. Hence, a fusion of both approaches might be beneficial for the overall performance of the morph detection system.

Table 2: D-EER performance comparison of proposed PRNU variance analysis based detector ($P_{en}|S_{disp}$) with baseline ($P_{pos}|S_{mean}$) proposed in [2]. The column *ALL* reports the D-EER including all attacks (*Morph* to *SHRP*).

		D-EER					
Algorithm	**Cells**	**Morph**	**EQU**	**SCL$_{50}$**	**SCL$_{75}$**	**SHRP**	**ALL**
$P_{pos}\|S_{mean}$	8	2.2%	33.8%	0.7%	0.0%	10.8%	15.7%
$P_{en}\|S_{disp}$	10	11.0%	15.9%	2.6%	3.8%	1.5%	10.5%
Difference		+ 8.8%	-17,9%	+1,9%	+3,8%	-9,3%	-5,2%

6. Conclusion and Future Work

When infiltrated during the enrolment process of a face recognition system, morphed face images pose a serious security risk, in particular in the context of ABC. In this work, a morph detector, which analyses the variance of PRNU-based features across image cells, is proposed. In contrast to related work [2], the presented approach is shown to be robust to diverse image post-processing techniques and even improves the D-EER for all investigated attacks, which include unaltered morphed images and various post-processing techniques, to 10.5%.

Compared to many other schemes, the presented system is expected to achieve high robustness, as it analyses relative changes of PRNU-based features across images regions rather than distinct texture features. Such changes inevitably occur if image morphing is applied. In order to avoid artefacts, some morphing algorithms paste morphed face regions within the convex hull of averaged landmarks into the outer region of one of the contributing face images. This would cause an even higher variance of PRNU features across image regions resulting in improved detection performance.

Future work will be focused on a more thorough analysis of the proposed approach, i.e. detection performance will be evaluated for bona fide and morphed images created from different face image databases using different morph generation algorithms. A comparison of the presented system against published face morph detectors will also be performed in future work. Finally, the creation of a database of printed and scanned (morphed) face images and a corresponding evaluation of the presented morph detection methods in different scenarios is subject to future work.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the Center for Research in Security and Privacy (CRISP). Furthermore, this work was also supported by the Austrian Science Fund (FWF) under Project No. P26630.

References

[1] FRONTEX – Research and Development Unit: Best practice technical guidelines for automated border control (ABC) systems, 2012. Version 2.0.

[2] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch. PRNU-based detection of morphed face images. In *2018 6th Intl. Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2018.

[3] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *Proc. Int. Joint Conf. on Biometrics (IJCB)*, 2014.

[4] M. Ferrara, A. Franco, and D. Maltoni. On the effects of image alterations on face recognition accuracy. In T. Bourlai, editor, *Face Recognition Across the Imaging Spectrum*. Springer International Publishing, 2016.

[5] M. Ferrara, A. Franco, and D. Maltoni. Face demorphing. *IEEE Transactions on Information Forensics and Security*, 13(4), 2018.

[6] M. Ferrara, A. Franco, and D. Maltoni. Face demorphing in the presence of facial appearance variations. In *2018 26th European Signal Processing Conf. (EUSIPCO)*, 2018.

[7] J. Fridrich. Digital image forensic using sensor noise. *IEEE Signal Processing Magazine*, 26(2), 2009.

[8] J. Fridrich. Sensor defects in digital image forensics. In H. Sencar and N. Memon, editors, *Digital Image Forensics: There is more to a picture than meets the eye*, chapter 6. Springer Verlag, 2012.

[9] M. Goljan, J. Fridrich, and J. Lukas. Camera identification from printed images. In *Proc. of SPIE, Electronic Imaging, Forensics, Security, Steganography, and Watermarking of Multimedia Contents X*. SPIE, 2008.

[10] M. Hildebrandt, T. Neubert, A. Makrushin, and J. Dittmann. Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps. In *2017 5th Intl. Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2017.

[11] ICAO. *ICAO Doc 9303, Machine Readable Travel Documents – Part 9: Deployment of Biometric Identification and Electronic Storage of Data in MRTDs (7th edition)*, 2015.

[12] International Organization for Standardization. Information technology – Biometric data interchange formats – Part 5: Face image data. ISO/IEC 19794-5:2005 consolidated, JTC 1/SC 37, 2005.

[13] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC IS 30107-3:2017, IT – Biometric presentation attack detection – Part 3: Testing and Reporting*.

[14] S. Jassim and A. Asaad. Automatic detection of image morphing by topology-based analysis. In *2018 26th European Signal Processing Conf. (EUSIPCO)*, 2018.

[15] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 2009.

[16] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann. Modeling attacks on photo-ID documents and applying media forensics for the detection of facial morphing. In *Proc. of the 5th ACM Workshop on Information Hiding and Multimedia Security - IHMMSec*. ACM Press, 2017.

[17] S. Z. Li and A. K. Jain. *Handbook of Face Recognition (2nd edition)*. Springer, 2011.

[18] X. Lin and C.-T. Li. Enhancing sensor pattern noise via filtering distortion removal. *IEEE Signal Processing Letters*, 23(3), 2016.

[19] A. Makrushin, T. Neubert, and J. Dittmann. Automatic generation and detection of visually faultless facial morphs. In *Proc. of the 12th Intl. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2017.

[20] A. Makrushin and A. Wolf. An overview of recent advances in assessing and mitigating the face morphing attack. In *2018 26th European Signal Processing Conf. (EUSIPCO)*, 2018.

[21] M. Mihcak, I. Kozintsev, and K. Ramchandran. Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising. In *Proc. of the 1999 IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, ICASSP '99*. IEEE, 2009.

[22] A. Mohammadi, S. Bhattacharjee, and S. Marcel. Deeply vulnerable: a study of the robustness of face recognition to presentation attacks. *IET Biometrics*, 7(1), 2018.

[23] T. Neubert. Face morphing detection: An approach based on image degradation analysis. In *Digital Forensics and Watermarking*. Springer International Publishing, 2017.

[24] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *Proc. of the British Machine Vision Conf. 2015, BMVC*, 2015.

[25] R. Ramachandra, K. B. Raja, and C. Busch. Detecting morphed face images. In *2016 IEEE 8th Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2016.

[26] R. Ramachandra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-CNN features for detecting digital and print-scanned morphed face images. In *2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017.

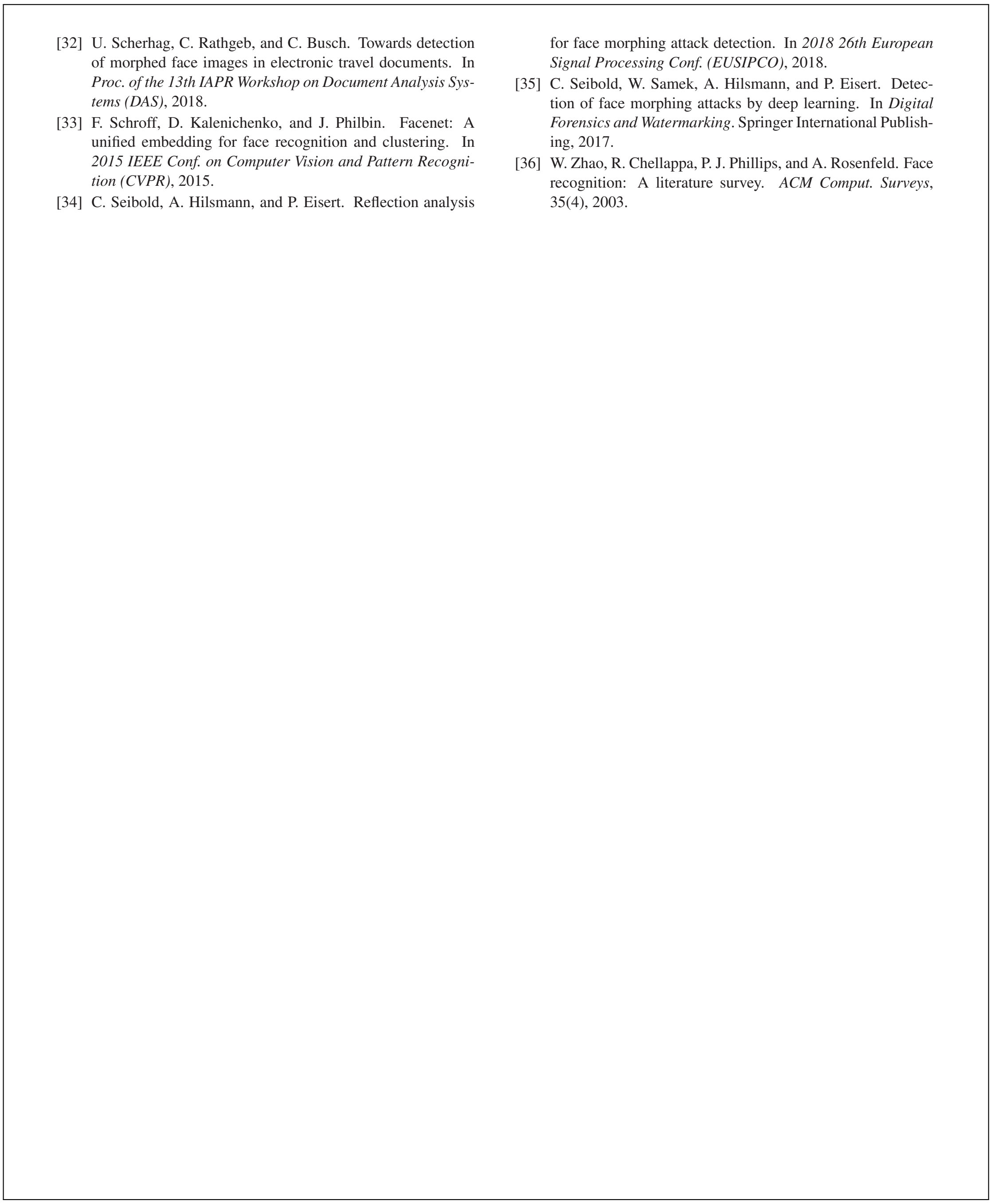
[27] U. Scherhag, D. Budhrani, M. Gomez-Barrero, and C. Busch. Detecting morphed face images using facial landmarks. In *Lecture Notes in Computer Science*. Springer International Publishing, 2018.

[28] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Raghavendra, and C. Busch. Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2017.

[29] U. Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch. On the vulnerability of face recognition systems towards morphed face attacks. In *2017 5th Intl. Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2017.

[30] U. Scherhag, C. Rathgeb, and C. Busch. Morph detection from single face images: a multi-algorithm fusion approach. In *Proc. of the 2018 Intl. Conf. on Biometrics Engineering and Application (ICBEA)*. ACM, 2018.

[31] U. Scherhag, C. Rathgeb, and C. Busch. Performance variation of morphed face image detection algorithms across different datasets. In *2018 6th Intl. Workshop on Biometrics and Forensics (IWBF)*. 2018.



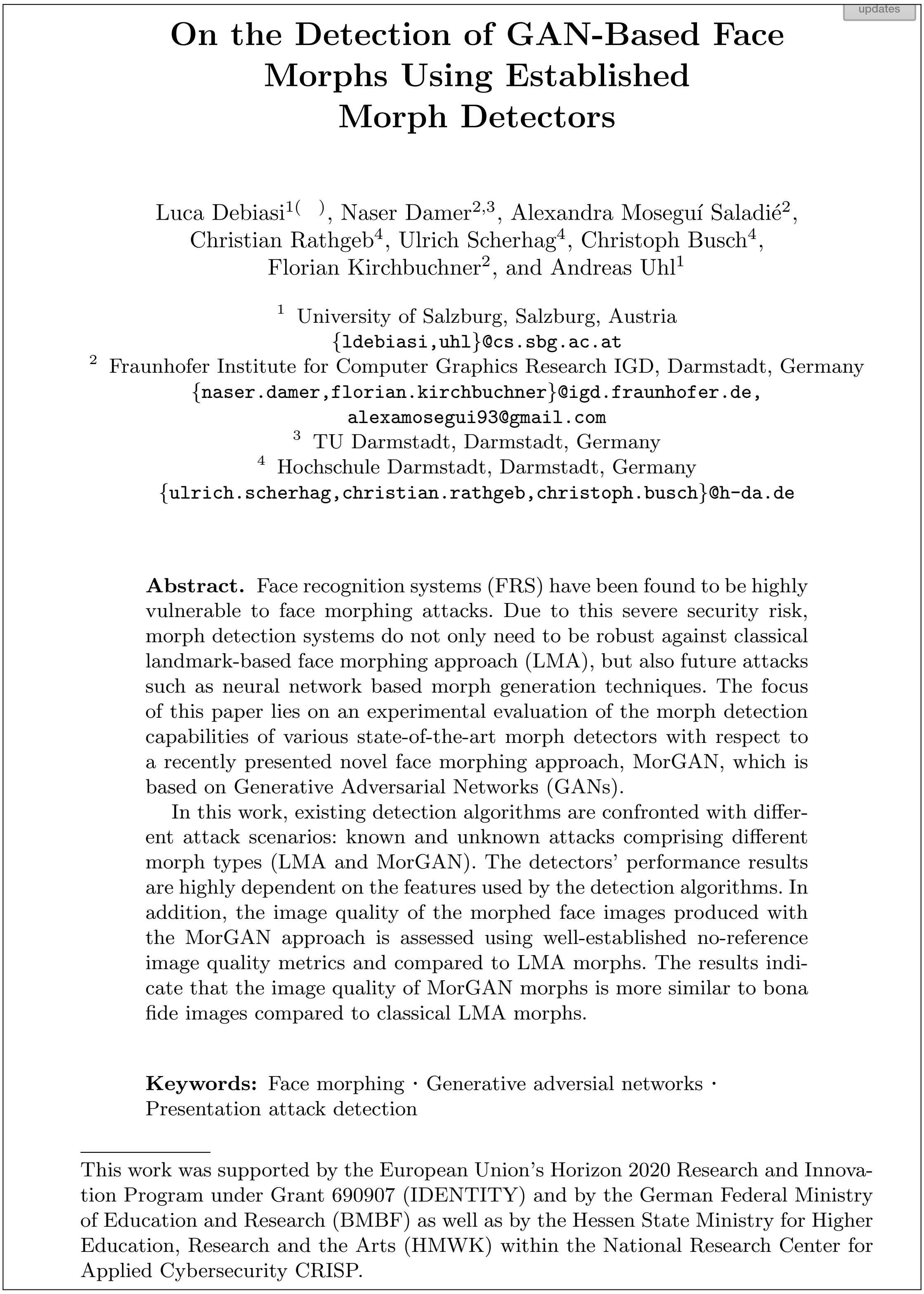
[32] U. Scherhag, C. Rathgeb, and C. Busch. Towards detection of morphed face images in electronic travel documents. In *Proc. of the 13th IAPR Workshop on Document Analysis Systems (DAS)*, 2018.

[33] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[34] C. Seibold, A. Hilsmann, and P. Eisert. Reflection analysis for face morphing attack detection. In *2018 26th European Signal Processing Conf. (EUSIPCO)*, 2018.

[35] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert. Detection of face morphing attacks by deep learning. In *Digital Forensics and Watermarking*. Springer International Publishing, 2017.

[36] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surveys*, 35(4), 2003.



On the Detection of GAN-Based Face Morphs Using Established Morph Detectors

Luca Debiasi[1]( ), Naser Damer[2,3], Alexandra Moseguí Saladié[2],
Christian Rathgeb[4], Ulrich Scherhag[4], Christoph Busch[4],
Florian Kirchbuchner[2], and Andreas Uhl[1]

[1] University of Salzburg, Salzburg, Austria
{ldebiasi,uhl}@cs.sbg.ac.at
[2] Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
{naser.damer,florian.kirchbuchner}@igd.fraunhofer.de,
alexamosegui93@gmail.com
[3] TU Darmstadt, Darmstadt, Germany
[4] Hochschule Darmstadt, Darmstadt, Germany
{ulrich.scherhag,christian.rathgeb,christoph.busch}@h-da.de

Abstract. Face recognition systems (FRS) have been found to be highly vulnerable to face morphing attacks. Due to this severe security risk, morph detection systems do not only need to be robust against classical landmark-based face morphing approach (LMA), but also future attacks such as neural network based morph generation techniques. The focus of this paper lies on an experimental evaluation of the morph detection capabilities of various state-of-the-art morph detectors with respect to a recently presented novel face morphing approach, MorGAN, which is based on Generative Adversarial Networks (GANs).

In this work, existing detection algorithms are confronted with different attack scenarios: known and unknown attacks comprising different morph types (LMA and MorGAN). The detectors' performance results are highly dependent on the features used by the detection algorithms. In addition, the image quality of the morphed face images produced with the MorGAN approach is assessed using well-established no-reference image quality metrics and compared to LMA morphs. The results indicate that the image quality of MorGAN morphs is more similar to bona fide images compared to classical LMA morphs.

Keywords: Face morphing · Generative adversial networks · Presentation attack detection

This work was supported by the European Union's Horizon 2020 Research and Innovation Program under Grant 690907 (IDENTITY) and by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the National Research Center for Applied Cybersecurity CRISP.

1 Introduction

Recently, automated face recognition systems (FRSs) are increasingly being used in different application scenarios, such as mobile device authentication or Automated Border Control (ABC). This wide spread deployment makes them attractive for attacks. In particular, their expected robustness to different environmental and user-specific conditions, e.g. varying illumination and subject poses, and the widespread use of deep neural networks in FRS has been found to increase their vulnerability against presentation attacks [14]. In this context, face morphing attacks have attracted notable interest from the research community in the recent past.

Ferrara *et al.* [6] unleashed the vulnerability of FRSs against attacks based on morphed face images, which can be introduced in the issuance process of electronic travel documents due to security gaps. They compared morphed images with images of the original subjects using two commercial face recognition solutions, and concluded with the high vulnerability of face recognition to such attacks. Further studies considered the human expert vulnerability to morphed face images when comparing faces [7, 20]. They found out that human experts fails most of the times in detecting morphing attacks.

Different solutions were developed to detect face morphing attacks. Ramachandra *et al.* [19] were first to propose the automated detection of morphed face images. They applied local image descriptors such as the Binarised Statistical Image Features (BSIF) that capture textural properties of the image, which are later classified using a Support Vector Machine (SVM). Later works looked into using convolutional neural network(CNN) based features [18], image quality measures [16], the effect of printing and re-scanning the images [23], and differences between triangulating and averaging the facial landmarks on the detection [17]. Recent works by Debiasi *et al.* [4] propose to exploit the Photo Response Non-Uniformity (PRNU) of an image sensor to detect morphed face images, which is a widely used tool in the field of Digital Image Forensics (e.g. image forgery detection).

A standardised manner to evaluate the vulnerability of biometric systems to morphing attacks was recently proposed by Scherhag *et al.* [22]. A recent work by Ferrara *et al.* [8] viewed the morphing attack detection problem from a different perspective by proposing an approach to revert the morphed face image (demorph) enough to reveal the identity of the legitimate document owner, given a bona fide capture.

Other works considered that it might be possible in practice to use a live probe image along with the investigated image to detect a morphing attacks. This was done either by looking at the differential vector between both images [24], analysing the absolute distances and angles of the landmarks in both images [21], analysing the directed distances between these landmarks [1], or using the live probe image for demorphing [8]. The mentioned works so far developed and evaluated their approaches based on morphing attacks databases that were created based on facial landmarks.

Recently, a work by Damer *et al.* [2] proposed a new possibility of morphing attacks. They built their solution on generative adversarial networks (MorGAN). They morphed the latent representation of the morphed images and generated the morphing attacks based on that morphed latent vector. These morphing attacks proved to be hard to detect in the cases where they were not considered in the training process of the morphing detector [2].

The work presented in this paper aims at evaluating the detectability of LMA- and GAN-based morphed face images in different attack scenarios (known and unknown attacks) using several state-of-the-art morph detectors based on different features. The experimental evaluation performed in this work gives a preliminary outlook on the detectability future face morphing attacks. These attacks might include novel morphing strategies such as GANs for face morph generation, where it is not clear how the morph detection performance is affected by the artefacts that they introduce. For example, it is not clear if the properties of the image's PRNU are preserved in morphed images generated using a GAN-based approach or if the properties are altered, which has a decisive impact on the detection performance of PRNU-based morph detection approaches. Furthermore, this work also includes an image quality assessment of morphed face images generated using the MorGAN approach compared to classical LMA morphs.

The paper is organised as follows: the MorGAN approach and data set are described in Sect. 2. The image quality assessment of the generated MorGAN images is reported in Sect. 3, while the experimental setup and investigated state-of-the-art morph detectors are described in Sect. 4. The experimental results are reported and discussed in Sect. 5 and the paper is concluded in Sect. 6.

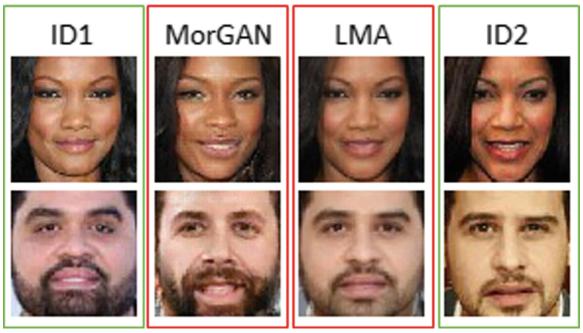


Fig. 1. Examples of the used morphing attacks, both the MorGAN and LMA. Original reference images are on the right and left.

2 MorGAN Dataset

A database containing attacks created by the conventional landmark-based morphing technique, as well as the recently MorGAN-based approach, is used in this work. This allows the evaluation of detection performance of known and unknown attacks of the investigated morph detection approaches.

The database is based on recent work by Damer *et al.* [2] foreseeing using GANs to create morphing attacks and built on the CelebA [12] data set.

The MorGAN database contains a total of 1500 bona fide references, 1500 bona fide probes, 1000 LMA morphing attacks, and 1000 MorGAN morphing attacks. The database is split into disjoint (identity and image) and equal train and test sets, each including 750 bona fide references, 750 bona fide probes, and 500 attack images from each of both attack types (LMA and GAN). Because of computational and structural limitations of the MorGAN approach, the MorGAN attack images are of 64×64 pixels size (below the ICAO recommendations). Examples of the resulting image attacks and the original images creating these attacks are presented in Fig. 1.

3 Quality of Morphed Face Images

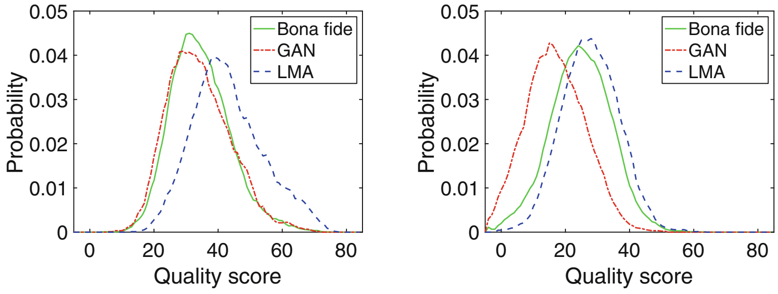
As shown in [2] by Damer *et al.*, the morphed face images contained in the MorGAN data set are capable of successfully attacking pre-trained FRS, i.e. OpenFace and VGG-Face. They conclude that MorGAN attacks are weaker than the LMA ones, however, still make successful attacks on both FRSs. It has to be noted that the MorGAN approach has only recently been presented and that images with higher quality and resolution are expected to be generated with future versions of the approach.

In this work, the insights on the vulnerability of FRSs against face morph presentation attacks are complemented by an image quality analysis of the MorGAN morphs, which is compared to the quality of bona fide images and LMA morphs. Ferrara *et al.* [6] demonstrated, that even human experts are not able to discriminate between bona fide and high quality morphed face images. Therefore, the image quality of morphed plays an important role, since common pattern recognition techniques and humans in particular can easily detect obvious artefacts within the images. For examples on such obvious artefacts, the reader is referred to [22]. In order to assess the image quality of the different images in the MorGAN data set (bona fide, MorGAN and LMA morphs), the following no-reference image quality metrics have been evaluated on all 1500 bona fide, 1000 MorGAN and 1000 LMA images: BIQI [15], BRISQUE [13], OG-IQA [10] and SSEQ [11]. To render a fair comparison with the MorGAN images possible, LMA and bona fide images have been downsized to the same resolution of 64×64 pixels. We did not consider any face-specific sample quality assessment metrics in this work due to the small resolution of the MorGAN images.

All image quality results are illustrated in Table 1, while only two selected quality metrics are presented in Fig. 2. Overall, the evaluation shows that the image quality of both morphed MorGAN and LMA images is very similar to the image quality of the bona fide images within the MorGAN data set. BIQI, OG-IQA and SSEQ show that the image quality score distributions of MorGAN images are more resemblant of the bona fide distribution compared to LMA

Table 1. Statistical properties of image quality metrics for bona fide images and LMA and MorGAN-based morphed images.

Metric	Property	Bona fide	MorGAN	LMA
BIQI	Mean	35.06	34.56	43.55
	Std	8.95	9.51	10.71
	Min	8.43	10.83	17.47
	Max	71.86	67.13	73.17
BRISQUE	Mean	25.22	17.23	28.30
	Std	9.13	9.45	8.50
	Min	−3.31	−12.71	2.28
	Max	59.76	90.29	59.29
OG-IQA	Mean	−0.82	−0.87	−0.74
	Std	0.09	0.07	0.10
	Min	−0.95	−0.95	−0.94
	Max	−0.25	−0.39	−0.39
SSEQ	Mean	30.25	29.51	37.80
	Std	9.30	7.82	7.70
	Min	−6.78	4.71	3.48
	Max	59.76	55.81	62.26



(a) BIQI

(b) BRISQUE

Fig. 2. Image quality score distributions of bona fide images compared to LMA and MorGAN-based morphs.

morphs. Only BRISQUE shows a different result, where the quality scores of LMA morphs are more alike the ones of bona fide images compared to MorGAN morphs. Due to time and space constraints, this deviation will be investigated more thoroughly in future work.

These results, using equally sized images of 64×64 pixels, reveal that morphed images generated with the MorGAN approach are more similar to bona fide images compared to the classical LMA approach in respect to their image quality, which is underlined by the distortion independence (BIQI), generalisability (OG-IQA) and closeness to human perception (SSEQ) of the image quality metrics supporting these results.

4 Experimental Setup

This study aims at investigating the detection performance of various morph detection approaches based on distinct features for MorGAN attacks. In particular, their ability of dealing with known and unknown attacks is of special interest, especially when future attacks based on unknown (neural network based) morphing techniques are considered.

4.1 Morph Detection Algorithms

Our morph attack detection methodology aims at enabling a wider range of conceptual evaluation and more diverse coverage of the state-of-the-art by considering image feature extraction methods of three different natures. One is the hand crafted classical image descriptors, the Local Binary Pattern Histogram (LBPH) [18], the second is based on transferable deep-CNN features [19] and the third type is based on the Photo Response Non-Uniformity (PRNU) [3,4]. All three types of features were previously utilised for the detection of face morphing attacks based on LMA approaches.

4.2 Experiments

The morph attack detection experiments are ordered by the feature type (CNN, LBPH, PRNU-VAR and PRNU-HIST) and by the type of attack, i.e. known or unknown and the type of morphs used for the attack (MorGAN and LMA). Due to the nature of the investigated detection algorithms and their design, the experiments had to be conducted in a slightly different manner for the various detectors, in order to ensure fair and comparable results. This has an effect on the sample size used for evaluation and the number of unknown attacks, which is described in more detail in the following.

Since CNN and LBPH are learning-based algorithms, the data is split into distinct train and test sets, both containing 750 bona fide images and 500 images for each attack type (LMA and MorGAN). A “known” attack (K) is given when the algorithm is evaluated with the same attack type as it is trained with, e.g. the algorithm was trained using LMA morphs and is evaluated on LMA morphs. An “unknown” attack (U), on the other hand, is given when different attack types are used to train and evaluate the algorithm, e.g. the algorithm is trained using LMA morphs and evaluated on MorGAN morphs. This leads to the following attack types for CNN and LBPH:

- K-LMA: Trained with LMA morphs, tested with LMA morphs.
- K-MorGAN: Trained with MorGAN morphs, tested with MorGAN morphs.
- U-LMA: Trained with MorGAN morphs, tested with LMA morphs.

– U-MorGAN: Trained with LMA morphs, tested with MorGAN morphs.

The two PRNU-based algorithms, PRNU-VAR and PRNU-HIST, do not rely on any training for classification, thus the whole data set, comprised of 1500 bona fide images and 1000 images for each attack type (LMA and MorGAN), is used for evaluation of the detectors. Therefore, all attacks with LMA or MorGAN morphs can be considered as "unknown" (U) for the PRNU-based algorithms. This leads to the following attack types for PRNU-VAR and PRNU-HIST:

– U-LMA: Tested with LMA morphs.
– U-MorGAN: Tested with MorGAN morphs.

4.3 Evaluation

The assessment of the morph detection performance is based on metrics defined in ISO/IEC 30107-3 [9]: Attack Presentation Classification Error Rate (APCER) and Bona Fide Presentation Classification Error Rate (BPCER), as suggested in literature [22]. APCER defines the proportion of morphed face presentations incorrectly classified as bona fide presentations, while BPCER is the proportion of bona fide presentations incorrectly classified as morphed face presentation attacks. The detection systems are evaluated at different operating points: The operation point of the system, where APCER = BPCER, is defined as detection equal error rate D-EER. Furthermore, two additional operation points, BPCER10 (where APCER = 10%) and BPCER20 (where APCER = 5%), are reported.

5 Morph Detection Results

The outcome of the morph detection experiments described in Sect. 4, are summarised in Table 2 and illustrated with DET plots in Fig. 3.

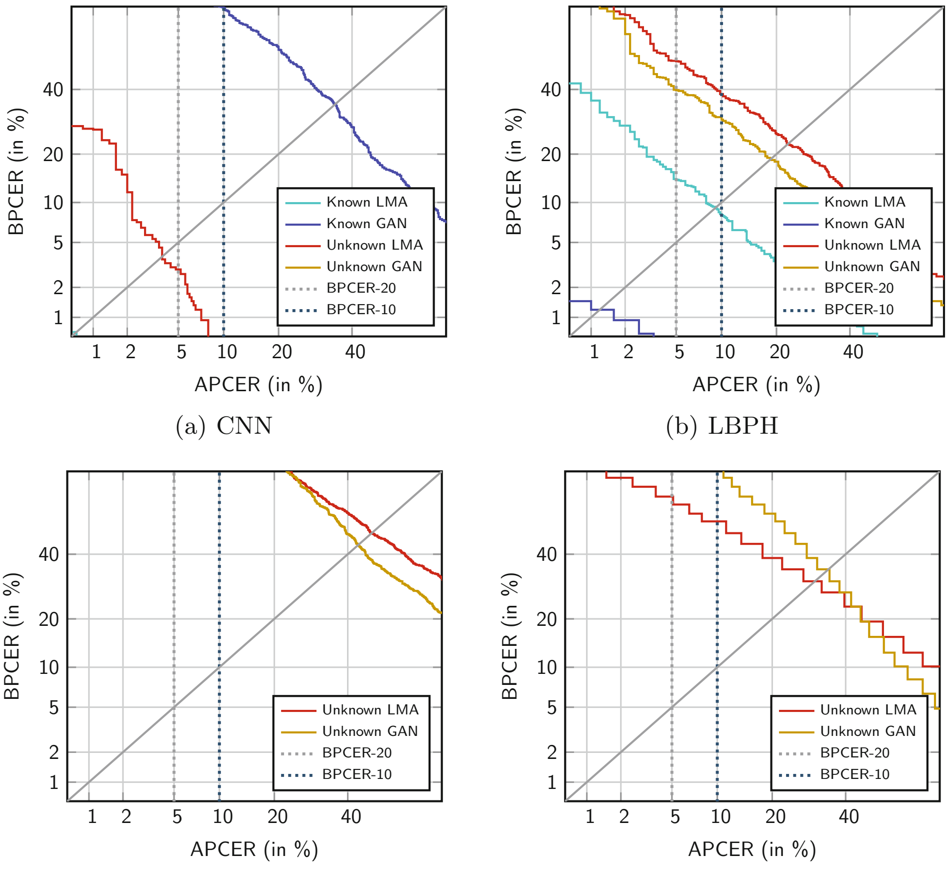
Table 2 shows the D-EER, BCPER10 and BCPER20 results for the various attack scenarios and morph detection algorithms described in Sect. 4. CNN shows the best performance at detecting LMA morphs, independent of the attacks being known or unknown. It achieves a perfect result for the K-LMA attack, and a D-EER of only 4% for U-LMA. However, it struggles in case of K-MorGAN or completely fails to detect U-MorGAN attacks. LBPH yields the overall lowest error rates among all morph detection algorithms and across all attack scenarios. It is able to detect both LMA and MorGAN morphs, but the performance gap between known and unknown attacks is very large. For known attacks, it is able to achieve low D-EERs of 9% for LMA and 1% for MorGAN attacks, while for unknown attacks the performance drops significantly to 23% and 19%, respectively. The results indicate that the CNN and LBPH detectors are not able to generalise well over different attack types, as it can be clearly seen in Fig. 3(a) and (b), which might be caused by the closed-set training design of both algorithms.

Table 2. Morph detection performance of investigated algorithms under different attack scenarios.

Algorithm	Attack type	D-EER	BCPER10	BCPER20
CNN	K-LMA	0.00	0.00	0.00
	K-MorGAN	0.34	0.67	0.78
	U-LMA	0.04	0.00	0.02
	U-MorGAN	0.50	0.90	0.95
LBPH	K-LMA	0.09	0.08	0.14
	K-MorGAN	0.01	0.00	0.00
	U-LMA	0.23	0.38	0.49
	U-MorGAN	0.19	0.29	0.39
PRNU-VAR	U-LMA	0.47	0.85	0.92
	U-MorGAN	0.43	0.85	0.92
PRNU-HIST	U-LMA	0.30	0.49	0.58
	U-MorGAN	0.33	0.69	0.81

The performance of the two PRNU-based algorithms is worse compared to the previously discussed CNN and LBPH algorithms, with D-EERs around 45% for PRNU-VAR and 30% for PRNU-HIST. Nonetheless, the results for these two algorithms show a very promising property: their stable performance across all attack types (known and unknown) and morph types (MorGAN and LMA). This consistency becomes evident when looking at Fig. 3(c) and (d). While they might not perform as well as CNN and LBPH in some cases, the results indicate a high potential for the generalisabilty of PRNU-based algorithms across different morph types, independently of the morph type being known or unknown. Furthermore, it can be observed that the PRNU of MorGAN morphs shows similar properties as the PRNU of LMA-based morphs, which leads to an almost equal detection performance for the PRNU-based detectors. Due to time and space constraints, a more thorough investigation of the PRNU signal resulting from the GAN operations is left for future research, in particular whether a PRNU-based identification of the source camera in images generated with GANs might still be possible. The D-EER performance of the two approaches is reported to be much better for larger images (320 × 320 pixels) in [4] and [3], thus we conclude that the overall poor performance for the PRNU-VAR and PRNU-HIST is a result of the small image size of 64 × 64 pixels in the MorGAN data set. It is commonly known in the field of Digital Image forensics, that the performance of PRNU-based approaches tends to degrade significantly with smaller image resolutions, as it is shown in [5].

Summarising the morph detection results, it can be observed that all investigated detection algorithms have their advantages and drawbacks. CNN works well for detecting LMA attacks, but fails at detecting MorGAN attacks. LBPH works quite well overall, but shows a high performance gap between known and unknown attacks, leaving it vulnerable for unknown attacks. PRNU-HIST and PRNU-VAR show an overall weak performance (presumably caused by the low image resolution), but they have the big advantage of being very stable across all evaluated attacks. If the general performance of the PRNU-based algorithms can be improved, it can be expected that they will show a high robustness against many unknown attack scenarios.



(a) CNN

(b) LBPH

(c) PRNU-VAR

(d) PRNU-HIST

Fig. 3. DET plots for investigated morphing detection algorithms and different attack scenarios.

6 Conclusion

The detection of morphed face images has become an important part of automated face recognition systems, due to their severe vulnerability to such attacks.

In this work, we investigate the performance of different state-of-the-art face morph detection algorithms on the recently proposed MorGAN data set. This data set, besides containing bona fide images and classical landmark-based morphs, also contains morphed images generated using the MorGAN approach. As the name implies, this novel type of morphed face images is created using Generative Adversarial Networks. The focus of this work lies on the evaluation of different attack scenarios: known and unknown attacks as well as different morph types. Furthermore, we also compare the image quality of MorGAN images to LMA based morphs using different well-established no-reference image quality metrics to evaluate the quality of generated morphs. The experimental evaluation performed in this work gives a preliminary prospect at the detection of future face morphing attacks, which might make use of unknown, most likely neural network based, morph generation techniques.

Summarising, the image quality assessment shows that the quality of MorGAN face morphs is closer to the quality of bona fide images as compared to classical LMA morphs, which underlines the capabilities of the MorGAN morph generation approach.

The morph detection performance results for the state-of-the-art detectors show that CNN fails at detecting the MorGAN morphs, but excels at detecting the classical LMA morphs. LBPH can achieve a very low D-EER of 1% for MorGAN and 9% for LMA morphs, but only in the case of known attacks. However, the performance of LBPH lacks consistency when confronted with unknown attacks. The two PRNU-based algorithms show a weaker overall performance of around 30% in the best case for both MorGAN and LMA morphs, which is most likely caused by the small image resolution.

Clearly, the MorGAN approach needs to be enhanced and further developed to produce images with higher resolutions, i.e. ICAO compliant images. This would allow for a more comprehensible analysis of the detectability and quality of the generated morphed face images.

References

1. Damer, N., et al.: Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts. In: Brox, T., Bruhn, A., Fritz, M. (eds.) GCPR 2018. LNCS, vol. 11269, pp. 518–534. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12939-2_36
2. Damer, N., Saladie, A.M., Braun, A., Kuijper, A.: MorGAN: recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In: 9th IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS 2018). IEEE (2018)
3. Debiasi, L., Rathgeb, C., Scherhag, U., Uhl, A., Busch, C.: PRNU variance analysis for morphed face image detection. In: Proceedings of the IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2018), Los Angeles, California, USA, pp. 1–8, October 2018
4. Debiasi, L., Scherhag, U., Rathgeb, C., Uhl, A., Busch, C.: PRNU-based detection of morphed face images. In: 2018 International Workshop on Biometrics and Forensics (IWBF 2018), pp. 1–7, June 2018

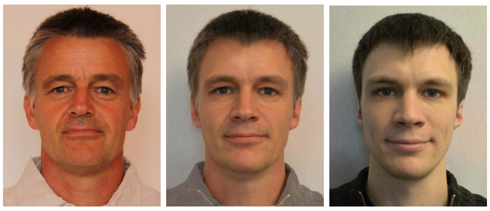
5. Debiasi, L., Uhl, A.: PRNU enhancement effects on biometric source sensor attribution. IET Biometrics **6**(4), 256–265 (2017)
6. Ferrara, M., Franco, A., Maltoni, D.: The magic passport. In: IEEE International Joint Conference on Biometrics (IJCB 2014), pp. 1–7. IEEE (2014)
7. Ferrara, M., Franco, A., Maltoni, D.: On the effects of image alterations on face recognition accuracy. In: Bourlai, T. (ed.) Face Recognition Across the Imaging Spectrum, pp. 195–222. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28501-6_9
8. Ferrara, M., Franco, A., Maltoni, D.: Face demorphing. IEEE Trans. Inf. Forensics Secur. **13**(4), 1008–1017 (2018)
9. ISO/IEC JTC1 SC37 Biometrics: ISO/IEC IS 30107–3:2017, IT - Biometric presentation attack detection - Part 3: Testing and Reporting
10. Liu, L., Hua, Y., Zhao, Q., Huang, H., Bovik, A.C.: Blind image quality assessment by relative gradient statistics and adaboosting neural network. Sig. Process. Image Commun. **40**, 1–15 (2016)
11. Liu, L., Liu, B., Huang, H., Bovik, A.C.: No-reference image quality assessment based on spatial and spectral entropies. Sig. Process. Image Commun. **29**(8), 856–863 (2014)
12. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV 2015), December 2015
13. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
14. Mohammadi, A., Bhattacharjee, S., Marcel, S.: Deeply vulnerable: a study of the robustness of face recognition to presentation attacks. IET Biometrics **7**(1), 15–26 (2018)
15. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal Process. Lett. **17**(5), 513–516 (2010)
16. Neubert, T.: Face morphing detection: an approach based on image degradation analysis. In: Kraetzer, C., Shi, Y.-Q., Dittmann, J., Kim, H.J. (eds.) IWDW 2017. LNCS, vol. 10431, pp. 93–106. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64185-0_8
17. Ramachandra, R., Raja, K.B., Venkatesh, S., Busch, C.: Face morphing versus face averaging: Vulnerability and detection. In: IEEE International Joint Conference on Biometrics (IJCB 2017), pp. 555–563. IEEE (2017)
18. Ramachandra, R., Raja, K.B., Venkatesh, S., Busch, C.: Transferable deep-CNN features for detecting digital and print-scanned morphed face images. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR, pp. 1822–1830. IEEE Computer Society (2017)
19. Ramachandra, R., Raja, K.B., Busch, C.: Detecting morphed face images. In: 8th IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS 2016), pp. 1–7. IEEE (2016)
20. Robertson, D.J., Kramer, R.S.S., Burton, A.M.: Fraudulent id using face morphs: experiments on human and automatic recognition. PLoS ONE **12**(3), 1–12 (2017)
21. Scherhag, U., Budhrani, D., Gomez-Barrero, M., Busch, C.: Detecting morphed face images using facial landmarks. In: Mansouri, A., El Moataz, A., Nouboud, F., Mammass, D. (eds.) ICISP 2018. LNCS, vol. 10884, pp. 444–452. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94211-7_48
22. Scherhag, U., et al.: Biometric systems under morphing attacks: assessment of morphing techniques and vulnerability reporting. In: International Conference of the Biometrics Special Interest Group (BIOSIG 2017), pp. 1–12 (2017)

Detection of Face Morphing Attacks Based on PRNU Analysis

Ulrich Scherhag, Luca Debiasi, Christian Rathgeb, Christoph Busch, and Andreas Uhl

***Abstract*—Recent research found that attacks based on morphed face images, i.e., morphing attacks, pose a severe security risk to face recognition systems. A reliable morphing attack detection from a single face image remains a research challenge since cameras and morphing techniques used by an attacker are unknown at the time of classification. These issues are commonly overseen while many researchers report encouraging detection performance for training and testing morphing attack detection schemes on images obtained from a single face database employing a single morphing algorithm. In this work, a morphing attack detection system based on the analysis of Photo Response Non-Uniformity (PRNU) is presented. More specifically, spatial and spectral features extracted from PRNU patterns across image cells are analyzed. Differences of these features for bona fide and morphed images are estimated during a threshold-selection stage using the Dresden image database which is specifically built for PRNU analysis in digital image forensics. Cross-database evaluations are then conducted employing an ICAO compliant subset of the FRGCv2 database and a Print-Scan database which is a printed and scanned version of said FRGCv2 subset. Bona fide and morphed face images are automatically generated employing four different morphing algorithms. The proposed PRNU-based morphing attack detector is shown to robustly distinguish bona fide and morphed face images achieving an average D-EER of 11.2% in the best configuration. In scenarios where image sources and morphing techniques are unknown, it is shown to significantly outperform other previously established morphing attack detectors. Finally, the limitations and potential of the approach are demonstrated on a dataset of printed and scanned bona fide and morphed face images.**

***Index Terms*—Biometrics, face recognition, face morphing, face morphing attack, morphing attack detection, photo response non-uniformity.**



(a) Subject 1 (b) Morph (c) Subject 2

Fig. 1. Example for a morphed face image (b) of subject 1 (a) and subject 2 (c) (images taken from [5]).

I. Introduction

FACE recognition systems have recently been exposed to be vulnerable against attacks based on morphed face images [1], [2]. Image morphing has been an active field of image processing research since the 1980s [3], [4] with a variety of application scenarios, especially in the film industry. Morphing techniques can be used to create artificial biometric samples that resemble the biometric information of two (or more) individuals in the image and feature domain. An example of a morphed face image is shown in Fig. 1. The morphed face image is successfully verified against probe samples of both subjects involved using state-of-the-art face recognition systems. This means that if a morphed face image is somehow stored as a reference in the database of a face recognition system, both individuals involved are successfully verified against this manipulated reference. Morphed face images thus pose a serious threat to face recognition systems, as the basic principle of biometrics, the unique link between the biometric reference data and the subject, is violated.

In many countries, the face image used for the ePassport application process is provided by the applicant either in analogue or digital form. In the scenario of a face morphing attack, a wanted *criminal* could morph his facial image with one of a lookalike *accomplice*. If the accomplice applies for an ePassport with the morphed face image, he will receive a valid ePassport equipped with corresponding document security features. It is important to note that morphed face images can be realistic enough to fool human examiners [6], [7] as well as commercial face recognition systems. Both the criminal and the accomplice could then be successfully verified against the morphed image stored in the ePassport. This means that the criminal can use the ePassport issued to the accomplice to pass through Automated Border Control (ABC) gates (or even human inspections at border crossings). The risk of this attack, called *face morphing attack*, is amplified by the fact that realistic face morphs can be generated by non-experts

Manuscript received March 7, 2019; revised July 5, 2019; accepted September 13, 2019. Date of publication September 23, 2019; date of current version October 30, 2019. This work was supported in part by the German Federal Ministry of Education and Research (BMBF), in part by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP (www.crisp-da.de), in part by the European Union's Internal Security Fund—Borders and Visa under Grant 842250-SOTAMD, and in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 690907 (IDENTITY). This article was recommended for publication by Associate Editor A. Ross upon evaluation of the reviewers' comments. *(Corresponding author: Ulrich Scherhag.)*

U. Scherhag, C. Rathgeb, and C. Busch are with the da/sec—Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany (e-mail: ulrich.scherhag@h-da.de).

L. Debiasi and A. Uhl are with the WaveLab—The Multimedia Signal Processing and Security Lab, Universität Salzburg, 5020 Salzburg, Austria.

Digital Object Identifier 10.1109/TBIOM.2019.2942395

using user-friendly face morphing software that is either freely available or can be purchased at a reasonable price.

In 2014 Ferrara *et al.* [1] were the first to thoroughly investigate the vulnerability of commercial face recognition systems to attacks based on morphed face images. So far, a considerable amount of morphing attack detection approaches has been published, see Section II. For a comprehensive survey the reader is referred to [2]. Proposed approaches can be categorized with respect to the considered morphing attack detection scenario:

- *No-reference morphing attack detection:* the detector processes a single image, e.g., the analysis of a printed image that is presented and scanned in a passport application procedure and subsequently stored in an electronic travel document or at any later point in time an off-line authenticity check of said document by police investigators (this scenario is also referred to as single image morphing attack detection or forensic morphing attack detection);
- *Differential morphing attack detection:* a trusted live capture from an authentication attempt serves as additional source of information for the morph detector, e.g., during authentication at an ABC gate (this scenario is also referred to as image pair-based morphing attack detection). Note that all information extracted by no-reference morph detectors might as well be leveraged within this scenario [8].

Obviously, the no-reference scenario turns out to be more challenging compared to the differential one. While the majority of no-reference approaches reports practical detection error rates, these are commonly evaluated on a dataset of bona fide and morphed face images which are extracted from a single (in-house) face database. In such an experimental setup the use of machine learning-based feature extractors or/and classifier increases the risk of overfitting, i.e., the robustness of morph detection algorithms may not be retained with regard to images stemming from other sources as shown in [9].

This work represents a significant extension of the preliminary studies towards PRNU-based morphing attack detection previously published in [5], [10]. The proposed system has been complemented by a more thorough investigation of different features and aggregation strategies, more specifically spatial features have been investigated in addition to spectral ones from previous work. Complementary to those efforts cross-database experiments on morphed face images generated by four different morphing algorithms have been conducted. The generalizability of the PRNU-based morphing attack detection across a wide range of distinct cameras of various makers is further investigated on a database specifically built for PRNU analysis in digital image forensics and it is shown that said database is suitable to determine the decision threshold for the proposed system. In addition, a database of printed and scanned face images is employed in evaluations. Moreover, in experiments the proposed system is benchmarked against state-of-the-art morphing attack detectors. Also, vulnerability analysis of the proposed concept with respect to potential attacks to circumvent the detection system is given.

The remainder of this work is organized as follows: related works are discussed in Section II. Fundamentals of PRNU extraction are explained in Section III. The proposed morph detection method is described in detail in Section IV. Experimental results are reported in Section V. Finally, conclusions are summarized in Section VI.

II. Related Work

In recent years, numerous no-reference face morphing attack detection schemes have been proposed. Published methods and their properties are summarized in Table I which has been derived from [2]. In some papers more than one system was presented, in such cases approaches that showed the best performance in detecting morphing attacks are listed. It is important to note that the generalizability/robustness of the published approaches could not be demonstrated. So far, there are no publicly accessible large databases of bona fide and morphed facial images and hardly any publicly available morph recognition algorithms which allow comprehensive experimental evaluations. The vast majority of published methods were trained and tested on various sequestered databases, which hampers reproducibility of results.[1] In addition, morph detection methods are usually trained and tested on a single database with a single morph generation algorithm. Based on these facts, a comparison of published approaches with respect to reported detection performance would be potentially misleading and is deliberately avoided in this work. However, it is expected that planned benchmark tests, e.g., by the National Institute of Standards and Technology (NIST) [40], will enable a meaningful quantitative comparison of published approaches in the near future.

Several researchers have suggested the use of general-purpose image descriptors, such as Local Binary Patterns (LBP) [41] or Binarized Statistical Image Features (BSIF) [42], which are widely used for biometric recognition. Ramachandra *et al.* [11] proposed a system based on a Support Vector Machine (SVM) trained on extracted BSIF features. For the training and evaluation of the SVMs, an internal database with morphed facial images was created. In a derivative version of the same database, Scherhag *et al.* [12] examined the accuracy of morphing detection on printed and scanned images using the proposed algorithm. Furthermore, Ramachandra *et al.* [13] proposed a Probabilistic Collaborative Representation Classifier (ProCRC) [43] trained on LBP features extracted from the color channels. The database used was an internal database derived from FRGCv2 [14]. The authors concentrate on the differences between morphed and averaged images in the evaluation.

A more complex method for morphing attack detection is proposed in [16], [17], where a Vietoris-Rips complex is formed from the reactions of uniform LBP extractors on the image. In [38] a high detection performance was shown by

[1] Also the morphed images used in this work can not be published due to licensing conditions as these are generated based on subsets of available image database collected by different institutions. However, efforts are currently made by different research laboratories to acquire new datasets of bona fide and morphed face images that shall serve future open benchmarks.

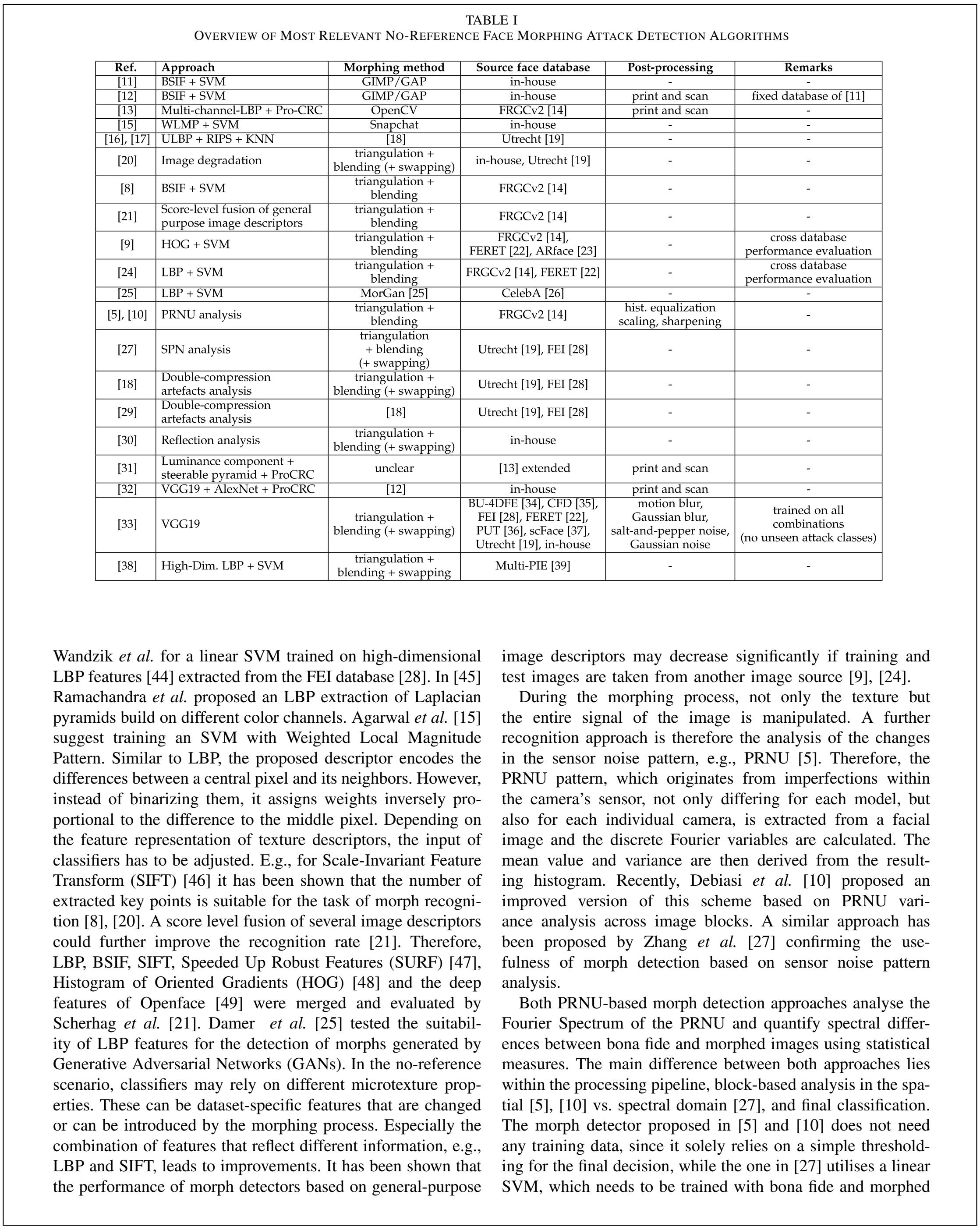


TABLE I
Overview of Most Relevant No-Reference Face Morphing Attack Detection Algorithms

Ref.	Approach	Morphing method	Source face database	Post-processing	Remarks
[11]	BSIF + SVM	GIMP/GAP	in-house	-	-
[12]	BSIF + SVM	GIMP/GAP	in-house	print and scan	fixed database of [11]
[13]	Multi-channel-LBP + Pro-CRC	OpenCV	FRGCv2 [14]	print and scan	-
[15]	WLMP + SVM	Snapchat	in-house	-	-
[16], [17]	ULBP + RIPS + KNN	[18]	Utrecht [19]	-	-
[20]	Image degradation	triangulation + blending (+ swapping)	in-house, Utrecht [19]	-	-
[8]	BSIF + SVM	triangulation + blending	FRGCv2 [14]	-	-
[21]	Score-level fusion of general purpose image descriptors	triangulation + blending	FRGCv2 [14]	-	-
[9]	HOG + SVM	triangulation + blending	FRGCv2 [14], FERET [22], ARface [23]	-	cross database performance evaluation
[24]	LBP + SVM	triangulation + blending	FRGCv2 [14], FERET [22]	-	cross database performance evaluation
[25]	LBP + SVM	MorGan [25]	CelebA [26]	-	-
[5], [10]	PRNU analysis	triangulation + blending	FRGCv2 [14]	hist. equalization scaling, sharpening	-
[27]	SPN analysis	triangulation + blending (+ swapping)	Utrecht [19], FEI [28]	-	-
[18]	Double-compression artefacts analysis	triangulation + blending (+ swapping)	Utrecht [19], FEI [28]	-	-
[29]	Double-compression artefacts analysis	[18]	Utrecht [19], FEI [28]	-	-
[30]	Reflection analysis	triangulation + blending (+ swapping)	in-house	-	-
[31]	Luminance component + steerable pyramid + ProCRC	unclear	[13] extended	print and scan	-
[32]	VGG19 + AlexNet + ProCRC	[12]	in-house	print and scan	-
[33]	VGG19	triangulation + blending (+ swapping)	BU-4DFE [34], CFD [35], FEI [28], FERET [22], PUT [36], scFace [37], Utrecht [19], in-house	motion blur, Gaussian blur, salt-and-pepper noise, Gaussian noise	trained on all combinations (no unseen attack classes)
[38]	High-Dim. LBP + SVM	triangulation + blending + swapping	Multi-PIE [39]	-	-

Wandzik *et al.* for a linear SVM trained on high-dimensional LBP features [44] extracted from the FEI database [28]. In [45] Ramachandra *et al.* proposed an LBP extraction of Laplacian pyramids build on different color channels. Agarwal *et al.* [15] suggest training an SVM with Weighted Local Magnitude Pattern. Similar to LBP, the proposed descriptor encodes the differences between a central pixel and its neighbors. However, instead of binarizing them, it assigns weights inversely proportional to the difference to the middle pixel. Depending on the feature representation of texture descriptors, the input of classifiers has to be adjusted. E.g., for Scale-Invariant Feature Transform (SIFT) [46] it has been shown that the number of extracted key points is suitable for the task of morph recognition [8], [20]. A score level fusion of several image descriptors could further improve the recognition rate [21]. Therefore, LBP, BSIF, SIFT, Speeded Up Robust Features (SURF) [47], Histogram of Oriented Gradients (HOG) [48] and the deep features of Openface [49] were merged and evaluated by Scherhag *et al.* [21]. Damer *et al.* [25] tested the suitability of LBP features for the detection of morphs generated by Generative Adversarial Networks (GANs). In the no-reference scenario, classifiers may rely on different microtexture properties. These can be dataset-specific features that are changed or can be introduced by the morphing process. Especially the combination of features that reflect different information, e.g., LBP and SIFT, leads to improvements. It has been shown that the performance of morph detectors based on general-purpose image descriptors may decrease significantly if training and test images are taken from another image source [9], [24].

During the morphing process, not only the texture but the entire signal of the image is manipulated. A further recognition approach is therefore the analysis of the changes in the sensor noise pattern, e.g., PRNU [5]. Therefore, the PRNU pattern, which originates from imperfections within the camera's sensor, not only differing for each model, but also for each individual camera, is extracted from a facial image and the discrete Fourier variables are calculated. The mean value and variance are then derived from the resulting histogram. Recently, Debiasi *et al.* [10] proposed an improved version of this scheme based on PRNU variance analysis across image blocks. A similar approach has been proposed by Zhang *et al.* [27] confirming the usefulness of morph detection based on sensor noise pattern analysis.

Both PRNU-based morph detection approaches analyse the Fourier Spectrum of the PRNU and quantify spectral differences between bona fide and morphed images using statistical measures. The main difference between both approaches lies within the processing pipeline, block-based analysis in the spatial [5], [10] vs. spectral domain [27], and final classification. The morph detector proposed in [5] and [10] does not need any training data, since it solely relies on a simple thresholding for the final decision, while the one in [27] utilises a linear SVM, which needs to be trained with bona fide and morphed images and makes the latter approach potentially more vulnerable against unknown morphing attacks. Furthermore, different PRNU extraction and enhancement techniques are used for both approaches. In contrast to [5], [10], the authors of [27] did not consider image post-processings. Also, no cross-database performance evaluations were performed.

Morphing attack detection methods based on continuous image degradation were proposed in [20], [50], [51]. The basic idea behind these methods is to continuously deteriorate the image quality, e.g., by JPEG compression, in order to generate several artificial self-references of a facial image. The distances between these references and the original image are then analyzed for morph detection. Ramachandra *et al.* [31] suggests the analysis of high frequencies. In their approach images are converted to grayscale and a controllable pyramid is built and a Collaborative Representation Classifier (CRC) is trained on the high frequencies. The database used was printed and scanned. An alternative to handcrafted feature extractors is the use of statistical machine learning on the unprocessed image to distinguish between morphed and bona fide images. Ramachandra *et al.* [32] suggested adapting two convolutional neural networks (CNNs) (VGG19 [52] and AlexNet [53]) by transfer learning and combining the intermediate features to train a CRC. In [54] three CNNs, namely VGG19, AlexNet and GoogLeNet [55], are assessed as pre-trained and non-pre-trained models with respect to their morph detection abilities. Also with these methods there is a potential problem of over-fitting. In particular, the resulting classifiers may prefer image sites where artefacts, such as shadows around the iris region, may occur due to an imperfect automated morphing process. In order to avoid over-fitting, Seibold *et al.* [33] trained a VGG19 network on a series of different images with two different databases, morphing algorithms and postprocessings (motion blur, Gaussian blur, salt and pepper noise, Gaussian noise). Since the CNN has been trained on all types of databases, morphing algorithms, and postprocessing, it is difficult to assess the resulting robustness of the classifier. Wandzik *et al.* [38] suggested to use pre-trained facial recognition networks, e.g., VGG-Face [56] or FaceNet [57], to detect morphing attacks. The high-level features generated by the networks are classified with a linear SVM.

Different approaches based on media forensics were presented, too. In several papers the detection of JPEG double compression artefacts for the purpose of morph detection was proposed [18], [29]. However, the presence of such artefacts implies a strong assumption of the image format of facial images used for morphing and the resulting morphed facial image. ICAO proposes to store facial image data in accordance with the specifications of the International Standard ISO/IEC 19794-5 [58]. More specifically, ICAO requires facial images to be stored in electronic travel documents with an average compressed size of 15kB to 20kB in JPEG or JPEG 2000 format [59], [60]. However, JPEG 2000 is the de-facto standard for electronic travel documents as it maintains a higher quality when compressing facial images to 15kB. Therefore, depending on the image size and the compression algorithm used, JPEG double compression artefacts may not be detected. A morph detection method based on reflection analysis in facial images is introduced by Seibold *et al.* [30]. The flash direction is estimated based on reflections detected in the eyes of a potentially morphed image. Reflections from the nose of the face are then analyzed. However, the ISO/IEC standard requires the absence of hot spots and reflections in facial images used in electronic travel documents. In particular, diffuse lighting, multiple symmetrical sources or other lighting methods should be used, i.e., a single bright “point” light source such as a camera-internal flash is not acceptable for imaging [58].

Apart from no-reference approaches differential morphing attack detection schemes have been presented, too. Most notably, face de-morphing [61], [62] and facial landmark-based approaches have been introduced [63], [64]. Additionally, some no-reference approaches, e.g., general-purpose image descriptors, can be extended to a differential scenario by estimating differences between feature vectors extracted from trusted live captures and potential morphs [8].

III. PRNU-Based Image Forensics

The photo response non-uniformity (PRNU), also known as sensor noise, has previously been utilised as a reliable tool to perform various forensic tasks such as device identification, device linking, recovery of processing history and the detection of digital forgeries. The PRNU origins from slight variations among individual pixels during the photoelectric conversion in digital image sensors. All digital image sensors cast this weak noise-like signal into all acquired images. Thus, the PRNU can be considered as an intrinsic property of all digital imaging sensors and an inherent part of their output.

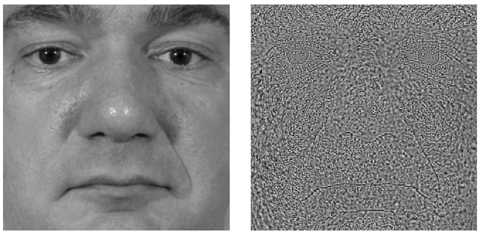
A. PRNU Extraction and Analysis

In this work, we make use of the PRNU to detect morphed face images. This systemic and individual pattern can be seen as an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering. The extraction of the PRNU noise residual from an image can be performed by applying Fridrich’s approach [65]. For each image I the noise residual W_I is estimated as described in Eq. (1),

$$W_I = I - F(I) \tag{1}$$

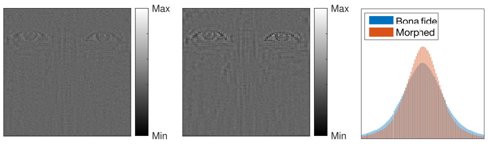
where F is a denoising function which filters out the sensor pattern noise. The extraction is performed using the denoising filter proposed by Mihcak *et al.* [66]. For further details on the denoising filter, we refer to [66]. Fig. 2 presents the extracted PRNU for an exemplary image. Further visualizations of PRNU signals extracted from face images can be found in [5], [10].

Since the PRNU extraction is relying on a denoising of the image, the resulting pattern might be contaminated with different signals, such as other high frequency image components, e.g., edges, or different types of non-unique artefacts (NUAs) [67]. Many alternative PRNU extraction schemes [68], [69], [70], [71], [72], [73], [74] and PRNU enhancements [75], [76], [77], [78], [79] have been



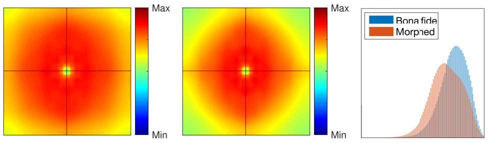
(a) Original image (b) Extracted PRNU

Fig. 2. PRNU extraction example for a pre-processed face image.



(a) Bona fide (b) Morphed (c) Histograms

Fig. 3. PRNU values and histograms of the PRNU extracted from a single bona fide image (a) and morphed face images (b). The PRNU values have been averaged over 500 randomly selected images of the FRGCv2 dataset.



(a) Bona fide (b) Morphed (c) Histograms

Fig. 4. DFT magnitude spectra and histograms of the PRNU extracted from bona fide and morphed face images. The DFT spectra have been averaged over 500 randomly selected images of the FRGCv2 dataset.

proposed in literature to attenuate different types of PRNU contaminations and improve the quality of the extracted PRNU in source camera identification scenarios. However, to the best of our knowledge, their impact on the general properties of the PRNU signal has not yet been extensively investigated. Therefore, we decided to rely on Mihcak *et al.*'s [66] denoising filter for the PRNU extraction.

The following essential properties, based on the characteristics of the PRNU described by Fridrich in [80], make the PRNU well suited for a face morph detection scenario:

1) *Dimensionality:* The sensor fingerprint is stochastic in nature and has a large information content, which makes it unique to each sensor.
2) *Unavoidability:* All imaging sensors exhibit PRNU.
3) *Universality:* The sensor fingerprint is present in every picture independently of the camera optics, camera settings, or scene content, with the exception of completely dark images.
4) *Permanence:* It is stable in time and under a wide range of environmental conditions (temperature, humidity, etc.).
5) *Robustness:* It survives lossy compression, filtering, gamma correction, and many other typical processing procedures. It is even reported to survive high quality printing and scanning [81].

Due to the criteria described above, the PRNU offers significant advantages over analysing other high-frequency image components to detect morphed face images.

According to Fridrich [65], the spectral characteristics of the PRNU reveal whether an image has been subject to further processing, e.g., non-geometrical operations have an influence on the strength of the embedded PRNU signal. Since the face morphing process involves non-linear warping and averaging operations, the distribution of the PRNU values is expected to change after these processing operations. Fig. 3 illustrates the PRNU and Fig. 4 the Discrete Fourier Transform (DFT) magnitude spectra obtained by averaging the extracted PRNU of 500 bona fide and 500 morphed face images from the FRGCv2 dataset, which is described in more detail in Section V.

These effects on the distribution of the PRNU values in the spatial domain can be observed in Fig. 3(c), where the distribution of morphed images is squashed compared to bona fide ones, i.e., the values around the mean of the distribution become more frequent and the values around the tails of the distribution become less frequent which leads to a steeper slope. Furthermore, some undesired components of the PRNU, e.g., edges in the image content, are emphasised in the morphed images, as it can be observed in Fig. 3(b). These effects are caused by the averaging operations applied during the morphing process.

The magnitude spectra of bona fide and morphed face images in Fig. 4, representing the frequency domain of the PRNU, show a clearly visible discrepancy among each other, where the most obvious differences can be observed in the reduction of high-frequency components within the morphed images' DFT magnitude spectrum as compared to the bona fide ones. Furthermore, the DFT spectrum of the morphed face images appears more compressed, i.e., the area covered by the large magnitudes is smaller compared to bona fide images.

These effects are caused by the previously mentioned operations involved in the face morphing process, which lead to changes in the distribution of the PRNU values. The approach presented in this work aims at exploiting these effects in order to perform a blind *no-reference* face morph detection.

B. Potential Attacks and PRNU Robustness

PRNU-based forensics and counter forensics can be considered as a cat-and-mouse game, since attacks and counter attacks are presented on a regular basis in the related literature. While attackers try to bypass various forensic approaches and conceal their counter-forensic approaches, techniques are developed to reveal such attacks.

The counter-forensic techniques proposed to overcome PRNU-based forensics can be divided into the following categories:

- *Destroying the Image Identity:* This class of counter forensic techniques tries to conceal the identity of an image and therefore prevents an identification of the image source or camera, respectively. Some examples are: removing the PRNU [82], [83], [84], [85], seam carving [86], [87], adaptive PRNU denoising [88]. Applying these techniques to morphed face images poses a lower threat to a PRNU-based morph detection system, since the aim is not to detect the image source, but to analyse the general properties of the PRNU signal. When the PRNU is destroyed, it can be assumed that its general properties are also not preserved.

- *Forging the Image Identity:* The goal of this class of counter forensic techniques is to fake the identity of an image, i.e., changing the identity of the image or concealing traces of its modification. Some examples for this are: Insertion of a differing PRNU signal [83], [85], fingerprint copy attacks [89], [90], [91], hiding of post-processing operations [83]. When applied to morphed face images, these type of counter forensic techniques can most likely be considered as a threat for a PRNU-based morph detection system, because their aim is to spoof an authentic image source, which usually contains similar characteristics to the PRNU of unaltered images. A potential attack on the PRNU-based morph detection system could involve extracting the PRNU from an authentic image and inserting it into a morphed image. This would restore the original properties of the PRNU when it is extracted again for the detection and therefore conceal the morphing operations.

Different approaches are proposed in literature to detect such intentional counter forensic attacks, e.g., the “Triangle Test” [92] and more recently Sameer *et al.* [93] proposed a deep learning based CNN model for the detection of counter forensic images. In biometrics, forging of the image identity has only been investigated for iris sensor data by Banerjee *et al.*, [94] and Uhl and Höller [95], where the detection of such attacks is furthermore evaluated in the latter.

Another type of attacks on the PRNU are unintentional ones, such as recompression, geometric transformations (cropping, scaling, rotation), photometric transformations and post-processing of the images. These attacks might occur unintentionally, i.e., when images are simply processed to enhance the appearance of a subject within the image, like it is often done for portrait photos. The PRNU has been shown to be resilient to photometric transformations [96] to a certain degree. While geometric transformations heavily affect the image source identification because they destroy the alignment of the PRNU signal, they are expected to not affect the general properties of the PRNU. However, post-processing of images, such as sharpening, blurring or contrast enhancement, can severely affect the PRNU. In previous work we showed that different post-processing techniques might even completely prevent a PRNU-based detection of morphed face images [5], [10]. Furthermore, recompression [97] is reported to alter the PRNU pattern after several passes in a way that source identification performance is affected. However, its influence on the general properties of the PRNU has not been investigated.

We consider intentional attacks on the PRNU to be less likely compared to unintentional ones, because the former require profound knowledge about the PRNU and its properties as well as an attacker with experience in the field. As the robustness of PRNU-based morph detection against simple post-processings has been already investigated in previous works [5], [10], an evaluation of four morphing algorithms has been included in order to provide a more comprehensive performance analysis in Section V-B. The four morphing algorithms picture a more realistic attack scenario, since they use different combinations of the simple post-processings. To address the question whether a PRNU-based approach can be applied for a wide range of distinct cameras, in Section V-C we evaluated the generalizability of the proposed morph detection approach on the Dresden Image Database [98] containing images from 63 different cameras from multiple manufacturers.

IV. Proposed System

Based on the observed effects of the face warping procedure on the spatial and spectral characteristics of the PRNU, in this work we propose a PRNU-based morph detection system which is able to discriminate between bona fide and morphed images. Therefore, we analyse the spatial and spectral characteristics of the PRNU in a *no-reference* manner, thus there is no need for a trusted bona fide reference image of one of the morphed subjects.

The proposed system relies on a divide and conquer principle and its processing steps are illustrated in Fig. 5. In the remainder of this section, we will discuss the various processing steps in more detail.

A. Preprocessing and PRNU Extraction

The first step of the system consists in extracting the facial region from a face image, which is normalised and then cropped to the facial area (320×320 pixels) before being converted to grayscale. This process is described in more detail in Section V-A.

Following, the PRNU is extracted from the preprocessed image, as described in Section III, using the wavelet-based denoising filter by Mihcak *et al.* [66] in conjunction with the filtering distortion removal (FDR) enhancement proposed in [79]. The extracted PRNU is then split into multiple equally sized cells. The proposed system is able to work with arbitrary splits from 1 cell (whole image) to N cells. In this work, only a cell size of 10×10 cells is investigated, because it yields the best performance according to previous work [5], [10]. In general, a larger number of cells is expected to further expose the non-linear transformations of the PRNU during the morphing process by putting stronger emphasis on local variations within an image. Eventually, we obtain N different cells $C_1, \ldots, C_N$. Fig. 5 shows an example of how the face image is preprocessed and the PRNU is extracted and split into 10×10 equisized cells.

B. Feature Extraction

The feature extraction is performed individually for each cell. In previous work [5], [10], only spectral features based on the DFT magnitude histogram and magnitude energy have

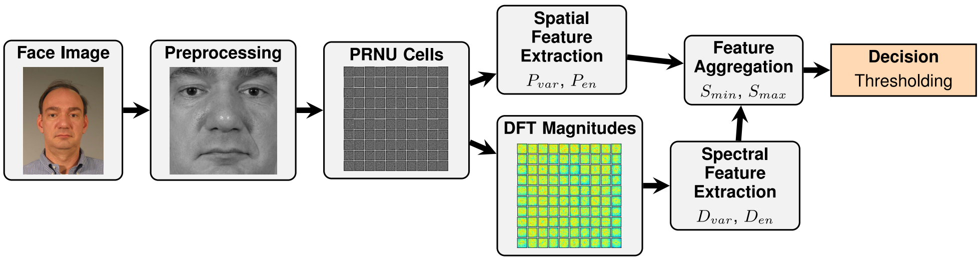


Fig. 5. Processing steps of the proposed PRNU-based morph detection system and different feature types: spatial features (upper path) and spectral features (lower path).

been investigated. In this work, two different feature types are investigated: spectral features based on the PRNU's DFT magnitudes and new spatial features based on the PRNU values, since the PRNU values are affected by the morphing procedures and post-processings in the spatial domain as well the spectral one.

Both feature types are described in more detail in the following.

1) Spatial Features: The newly proposed spatial features aim at analysing the distribution of the PRNU values, which is observed to differ between bona fide and morphed images according to Fig. 3(a) and Fig. 3(b).

For the first spatial feature, P_{var}, the histogram of the PRNU values is computed, which is constrained to a range of $[-5, 5]$ and divided into 100 bins. These values have been selected by analysing the DFT spectra of extracted PRNUs of bona fide and morphed images. Due to the different slope of bona fide and morphed image's PRNU value distributions that can be observed in Fig. 3(c), we decided to compute the variance of the histogram bin frequencies P_{var}, which we defined as

$$P_{var} = \frac{1}{B}\sum_{n=1}^{B}\left(H_P(n) - \bar{H_P}\right)^2 \tag{2}$$

where B is the number of bins in the PRNU cell's histogram H_P. $\bar{H_P}$ represents the mean frequency of the histogram bins.

As second spatial feature, we consider the energy of the PRNU values, P_{en}, which is defined as

$$P_{en} = \sum_{x \in V}|x|^2 \tag{3}$$

where x is a value within all PRNU values V of a cell.

As the Eqs. (2) and (3) show, both spatial features yield a simple scalar value SV for each PRNU cell.

2) Spectral Features: In order to compute the spectral features, the first step consists in obtaining the frequency spectrum of the PRNU in each cell, which is done by means of the DFT. The resulting magnitude spectrum, which is illustrated in Fig. 4(a) and Fig. 4(b) respectively, reveals the alterations of the PRNU signal caused by the morphing process.

These effects are quantified, on one hand, by calculating the DFT magnitude histogram to represent the magnitude distribution within the spectrum. As described in Section III, a shift of the magnitude distribution can be observed for morphed images. The DFT magnitude histograms are constrained to the same universal range of [0, 8] and are divided into 100 bins. These values have again been estimated by analysing the DFT spectra of extracted PRNUs of bona fide and morphed images. Based on the observations in Section III, we select the variance of the histogram D_{var} as being suited for the discrimination between bona fide and morphed images. We obtain D_{var} in a similar manner as the previously described P_{var}:

$$D_{var} = \frac{1}{B}\sum_{n=1}^{B}\left(H_M(n) - \bar{H_M}\right)^2 \tag{4}$$

where B is the number of bins in a cell's DFT magnitude histogram H_M, with $\bar{H_M}$ being the mean frequency of the histogram bins.

On the other hand, we propose to compute the energy of the PRNU's DFT magnitudes D_{en}, as defined in Eq. (5), where M are the DFT magnitudes within a cell and x their respective values.

$$D_{en} = \sum_{x \in M}|x|^2 \tag{5}$$

As for the spatial features, both spectral features yield a simple scalar value SV for each PRNU cell when considering Eqs. (4) and (5).

C. Feature Aggregation

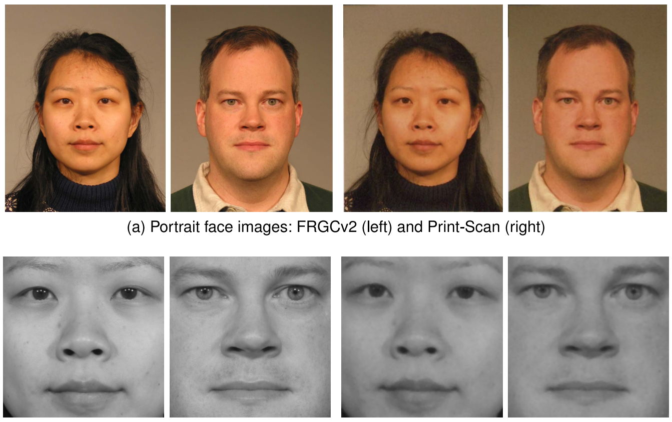
After obtaining the scalar values SV for all cells C_n, the values are aggregated to obtain a global aggregation score A for the image. We investigated various strategies, where we present the two best performing ones. The aggregation strategies used in this work are:

$$A_{min} = \min_{\forall n \in 1 \ldots N} SV_n \tag{6}$$

$$A_{max} = \max_{\forall n \in 1 \ldots N} SV_n \tag{7}$$

where N is the number of total cells and SV_n is the feature (scalar value) obtained for the cell C_n, as described in the previous processing step.

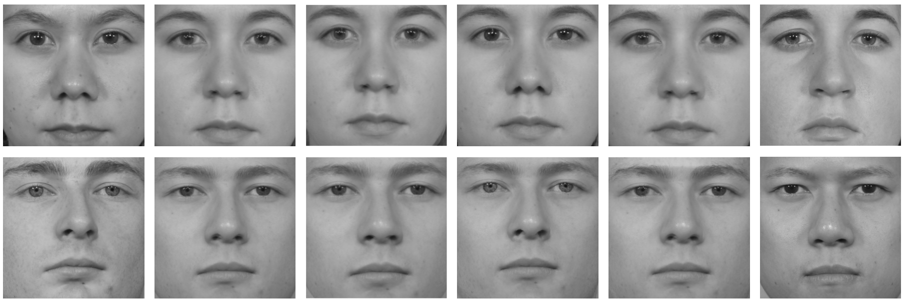
A_{min} yields the minimum score among the individual cells, while A_{max} characterizes maximum score among all cells. As already mentioned, we obtain a single scalar value A for each image using one of the Eqs. (6) or (7).



(a) Portrait face images: FRGCv2 (left) and Print-Scan (right)

(b) Pre-processed face images: FRGCv2 (left) and Print-Scan (right)

Fig. 6. Examples of bona fide portrait and pre-processed face images of the used datasets. Due to the printing and scanning face images from the Print-Scan dataset exhibit slightly lower resolution.



(a) Subject 1 (b) OpenCV/dlib (c) FaceMorpher (d) FaceFusion (e) UBO (f) Subject 2

Fig. 7. Used morphing algorithms applied to a female (top) and a male (bottom) image pair. Note that the FaceFusion algorithm uses the inner eye regions and nostrils of subject 1 in order to avoid artefacts in these regions.

D. Decision

The final decision, whether a face image has been created through morphing of multiple images or not, is taken by a simple thresholding.

Previous work [5] showed that a one dimensional decision was not able to reliably detect some of the post-processed morphed images for some spectral features. Hence, we introduce an additional decision step and derive a mean value $\bar{B}$ from bona fide images, where the characteristics of the PRNU are well known. With this property, we can calculate the distance D of an investigated image to bona fide images as

$$D = |A - \bar{B}| \tag{8}$$

$$\bar{B} = \frac{1}{N_B} \sum_{n=1}^{N_B} A \tag{9}$$

where A is the cell aggregation result, $\bar{B}$ is the mean variation of the N_B bona fide images.

It has to be noted, that this distance calculation is only applied for the two spatial and spectral energy-based features P_{en} and D_{en}, while it is not calculated for the histogram-based features P_{var} and D_{var}, due to the histogram-based features yielding more consistent scores among different post-processings which can be classified with a one dimensional threshold.

If the distance calculation is applied, the final decision for a presented face image is taken by thresholding the calculated distance D. Otherwise, the final decision simply relies on thresholding of the value A, which is obtained directly from the cell aggregation.

V. Experiments

In the following subsection the experimental setup, i.e., used databases, morphing algorithms, baseline systems and performance metrics, are described. Subsequently, the detection performance of the proposed systems and the baseline systems is reported and discussed. Further, the generalizability of the proposed PRNU-based morph detection approach with respect to utilized cameras and printed and scanned face images is investigated.

A. Experimental Setup

Performance evaluations are conducted based on a subset of 1,948 images selected from the FRGCv2 [14] face image database. Face images have been manually filtered to meet ICAO requirements for electronic travel documents [59], e.g., frontal pose, neutral expression, homogeneous background and sufficient resolution (at least 90 pixels between left and right eye center). Images of this database have been developed using a Fujifilm Frontier 5700 R Minlab and scanned using a Epson DS-50000 Scanner at 300 dpi to obtain the Print-Scan database of equal size. In addition, a subset of 1,058 images from the FERET [22] face image database which exhibit the same properties are used for training purposes of baseline morph detection algorithms. Note that the latter database is not used for evaluation of the proposed PRNU-based morph detection scheme since it has been acquired using an analog camera. PRNU is primary caused by Pixel Uniformity Noise related to the sensor which are non-existent if images are acquired with a film camera, i.e., only the PRNU signal of the sensor inside the scanner used to digitize the images might be present in this case. Instead, the Dresden Image Database [98] is used for training the PRNU-based morph detection schemes to underline the claim that the proposed PRNU-based morph detector is not dependent of a specific camera unit, since it contains images from 63 distinct cameras from various models and manufacturers. More details on how the bona fide and morphed images have been generated using the Dresden Image Database are given in Section V-C.

In a pre-processing step the face of a subject is segmented and normalized according to eye coordinates detected by the *dlib* landmark detector [99]. Subsequently, the normalized region is cropped to 320×320 pixels to ensure that the morph detection algorithm is only applied to the facial region. Finally, the cropped face part is converted to a grayscale image. Examples of original face images (cropped to portrait format) and pre-processed face images of the FRGCv2 and Print-Scan database are depicted in Fig. 6.

The subsets are split into images used for morph creation and images used as bona fide references. The resulting database constellation is listed in Table II. In order to generate a great variation of morphs, four morphing algorithms were employed:

1) *OpenCV/dlib:* a self-scripted morphing algorithm based on th "Face Morph Using OpenCV" tutorial[2] using the dlib landmark detector [99].
2) *FaceMorpher[3]:* an open-source implementation using python.
3) *FaceFusion[4]:* a proprietary morphing algorithm.
4) *UBO:* the morphing tool developed by the University of Bologna, as used, e.g., in [61].

TABLE II
Number of Subjects, Bona Fide and Morphed Face Images of Used Datasets. "F" and "M" Indicate Female and Male Subjects, Respectively

Database	Subjects	Face images	
		bona fide	morphed
FRGCv2	533 (231 F, 302 M)	984	964*
Print-Scan	533 (231 F, 302 M)	984	964*
FERET	529 (200 F, 329 M)	529	529*

*per morphing algorithm

In order to be able to conduct comparable experiments, the same combination of morphed face images was created for each of the listed algorithms. All algorithms detect corresponding landmarks in two face images to be morphed which are averaged. Subsequently, both face images are warped accordingly. Finally, alpha-blending is performed to create the morphed face image. All morphs were created in a way such that both used images tend to contribute equally to the inner facial region. Note that FaceFusion and UBO morphing algorithms are closed-source and might apply certain image post-processing methods to enhance the quality of resulting morphs. Examples of cropped facial regions of morphed face images generated all four morphing algorithms are shown in Fig. 7.

The vulnerability of a COTS facial recognition system to attacks based on the generated morphed face images is assessed by using the metrics specified in [100], in particular the Mated Morph Presentation Match Rate (MMPMR). This measure is an adaptation of the general Impostor Attack Presentation Match Rate (IAPMR) introduced in ISO/IEC 30107−3 [101] and is defined as the proportion of attack presentations using the same type of presentation attack instruments in which the target reference matches. In the adaptation, however, the MMPMR covers the fact that not one target subject (contained in the morphed reference) is matched - but for a successful face morphing attack, both data subjects that previously contributed to the morphed image are expected to match.

Using the default decision threshold of the COTS facial recognition system, an MMPMR of 1 is obtained across all used face image databases and morphing techniques. This means that all facial images of individuals contributing to a morphed facial image are successfully compared to it, so that the attacks have a 100% chance of success.

As baseline face morphing attack detection systems Local Binary Patterns (LBP) [102], Binarized Statistical Image Features (BSIF) [42], FaceNet features [57] and the FS-SPN analysis of [27] are applied. At feature extraction for LBP and BSIF the pre-processed face image is optionally divided into 4×4 cells to retain local information. That is, feature extractors are applied pixel-wise storing feature value in histograms for each texture cell. The final feature vector is formed as a concatenation of histograms extracted from each cell. While LBP simply processes neighboring pixel values of each pixel, BSIF utilizes specific filters learned from a set of

[2] http://www.learnopencv.com/face-morph-using-opencv-cpp-python/
[3] https://github.com/alyssaq/face_morpher
[4] http://www.wearemoment.com/FaceFusion/

TABLE III
PERFORMANCE RESULTS IN TERMS OF D-EER (IN %) FOR DIFFERENT CONFIGURATIONS OF THE BASELINE MORPHING ATTACK DETECTION SYSTEMS. BEST PERFORMING SYSTEMS ARE MARKED BOLD. μ IS THE MEAN ERROR AND σ^2 THE VARIANCE OVER ALL MORPHING METHODS

System	Morphing method				μ	σ^2
	OpenCV/dlib	FaceMorpher	FaceFusion	UBO		
LBP$_{1\times1}$ [8]	35.5%	15.3%	28.1%	26.1%	26.5%	±5.82
LBP$_{4\times4}$ [8]	20.5%	**4.2%**	14.7%	**12.7%**	**13.0%**	±4.66
BSIF$_{1\times1}$ [8]	27.6%	26.0%	16.7%	17.6%	22.0%	±4.87
BSIF$_{4\times4}$ [8]	27.4%	29.0%	**7.9%**	16.6%	20.2%	±8.57
FaceNet [57]	30.1%	29.8%	32.0%	33.2%	31.3%	±1.33
FS-SPN [27]	**17.5%**	4.9%	30.8%	19.5%	18.2%	±7.0

images. For details on these texture descriptors the reader is referred to [42], [102]. The use of these well-established general purpose texture descriptors has shown to be successful in diverse texture classification problems. As the process of image morphing is expected to cause changes in textual properties between bona fide and morphed face images said texture descriptors have been shown to reveal competitive morphing attack detection performance [8], [11], [12], [21]. Minimum filter sizes of 3×3 pixels which have been reported to reveal best detection performance in [8] are used for both texture descriptors. In the training stage feature vectors are extracted for each baseline system and SVMs with Radial Basis Function (RBF) kernels are trained to distinguish between bona fide and morphed face images. Similarly, an SVM is trained with deep facial features extracted from cropped face image using the FaceNet recognition system. This approach resembles the schemes proposed in [31], [33]. The SVM-based classifiers of these morph detection schemes are trained on the subset of the FERET image database. Eventually, the pre-trained open-source implementation[5] of [27] is directly applied for morph detection. The major advantage of the proposed PRNU-based morph detection over the baseline algorithms is that it does not need any training. Only for some of the proposed features, a pre-computed decision threshold has to be computed. In such cases, the threshold has been estimated on the Dresden image database [98].

The performance of the detection algorithms is reported according to metrics defined in ISO/IEC 30107−3 [101]. The Attack Presentation Classification Error Rate (APCER) is defined as the proportion of attack presentations using the same presentation attack instrument species incorrectly classified as bona fide presentations in a specific scenario. The Bona Fide Presentation Classification Error Rate (BPCER) is defined as the proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario. The D-EER, i.e., the operation point where APCER = BPCER, is used as general operation point and reported for the different morphing methods.

B. Performance Evaluation

Table III lists the D-EERs for different configurations of the baseline systems. It can be observed that morphs created using OpenCV with dlib are generally harder to detect, in contrast to the images created by other morphing algorithms.

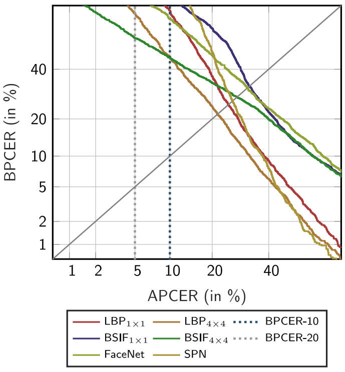
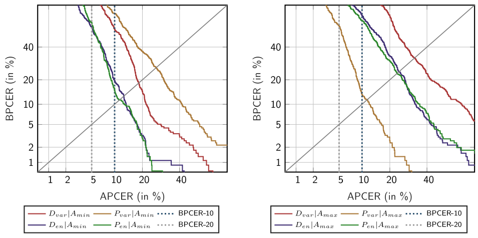


Fig. 8. DET curves for different configurations of the baseline morphing attack detection systems in the presence of all morphing attacks on FRGCv2.



(a) A_{min}

(b) A_{max}

Fig. 9. DET curves for different configurations of the baseline morphing attack detection system in the presence of all morphing attacks.

However, FS-SPN performs best detecting morphs created with OpenCV and dlib, but the detection rate drops when detecting morphs created by FaceFusion or the UBO algorithm. In contrast, BSIF$_{4\times4}$ shows improved performance for detecting FaceFusion morphs, but lacks detecting morphs created by OpenCV. The DET curves for the baseline systems in presence of all morphing attacks are shown in Fig. 8. In summary, it appears that a heterogeneous training and test database as well as the utilization of different morphing algorithms significantly deteriorate the detection performance of the baseline systems leading to significantly worse results to what has been reported in previous works.

Performance results for the proposed PRNU-based morphing attack detection scheme for best performing feature extractors and cell aggregation techniques are summarized in Table IV. DET plots for the best performing proposed approaches across all post-processings are shown in Fig. 9. In addition, Fig. 10 compares the average D-EERs and their variances of all proposed morphing attack detection schemes to the baseline systems. In contrast to the baseline systems, the PRNU-based approaches yield low error rates detecting morphs created using OpenCV and dlib, but struggle detecting FaceFusion morphs. However, compared to the baseline

[5] https://github.com/Le-BingZhang/FS-SPN

TABLE IV
PERFORMANCE RESULTS IN TERMS OF D-EER (IN %) FOR DIFFERENT CONFIGURATIONS OF THE PROPOSED PRNU-BASED MORPHING ATTACK DETECTION SYSTEMS. BEST PERFORMING SYSTEMS ARE MARKED BOLD. μ IS THE MEAN ERROR AND σ^2 THE VARIANCE OVER ALL MORPHING METHODS

Feature Extraction	Cell aggregation	Morphing method				μ	σ^2
		OpenCV/dlib	FaceMorpher	FaceFusion	UBO		
D_{var}	A_{min}	1.7%	7.9%	40.3%	15.8%	18.3%	± 2.86
	A_{max}	13.1%	20.2%	50.0%	29.6%	32.3%	± 2.56
P_{var}	A_{min}	15.8%	7.5%	45.9%	26.2%	26.3%	± 2.76
	A_{max}	**0.2%**	**0.5%**	**28.2%**	**8.9%**	**11.2%**	± **1.73**
D_{en}	A_{min}	0.6%	1.0%	29.5%	11.0%	12.4%	± 1.84
	A_{max}	7.5%	5.3%	47.9%	21.7%	24.1%	± 3.84
P_{en}	A_{min}	0.2%	0.6%	28.7%	9.6%	11.3%	± 1.79
	A_{max}	11.8%	4.8%	44.0%	22.6%	23.3%	± 2.92

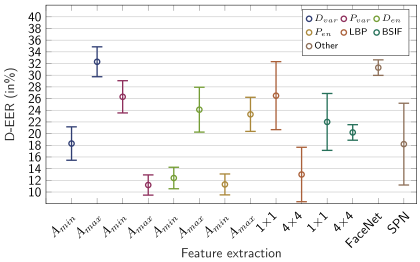


Fig. 10. Error bars of D-EERs for different configuration of the proposed PRNU-based morphing attack detection system and the baseline morphing attack detection systems in presence of all morphing attacks.

systems average D-EER are observably lower and exhibit smaller standard deviations. Additionally, smaller variance in detection performance across different datasets and morphing algorithms are obtained, which is vital for an application of any morphing attack detection algorithm in real world scenarios where said parameters are unknown.

Compared to the baseline systems, significantly improved results are achieved for the newly proposed spatial features, i.e., P_{var} followed by P_{en}, which significantly outperform the baseline systems. The spectral D_{en} feature, proposed in previous work, also obtains very competitive results on this new dataset. Another aspect to note is that the energy-based features D_{en} and P_{en}, whose mean bona fide threshold $\bar{B}$ has been determined on the Dresden image database, underlines the generalisability of the approach in regard to cameras from different models and manufacturers.

At this point, it is important to note that morphing attack detection algorithms analyze cropped faces only. Thereby higher generalizability is achieved since outer facial parts can be created in different ways during morph creation. Many morph generators copy the outer facial image part of one subject contributing to the morph, e.g., in [29], [61]. In such cases, the PRNU signal of the outer part of the morph is expected to remain almost unaltered. That is, if the proposed PRNU-based morphing attack detection schemes are extended to analyze the entire face image, a variance-based cell aggregation is expected to reveal improved results for detecting morphs created in the aforementioned way.

Overall, some of the proposed PRNU-based morphing attack detection configurations reveal promising results considering the challenging experimental setup. In contrast to trained morphing attack detection schemes, e.g., [32], [54], the proposed schemes do not rely on the presence of distinct artefacts, e.g., ghost artefacts, which might occur due to imperfect morph creation. Hence, similar results are to be expected if advanced morphing algorithms are developed which allow for an automated creation of morphs comprising less or no artefacts.

C. Generalizability Across Cameras

As mentioned in Section III, the proposed PRNU-based morph detection system relies on changes in the distribution of the PRNU values. Since the PRNU differs for each camera, it might contain camera (model) specific contaminations (non-unique artefacts) that might affect the PRNU values' distribution.

In order to investigate the generalizability of the proposed morph detection approach and due to a lack of suitable face image datasets acquired with different cameras, we decided to fall back to the Dresden image database [98], which offers images from multiple cameras and even multiple instances of the same camera model. More specifically, we selected the *flatfield* dataset, since it contains images beneficial for PRNU extraction, i.e., bright images of an evenly illuminated surface, which do not contain any contaminations from the image content like edges or other high-frequency patterns. The flatfield dataset contains images from 63 distinct digital cameras from 20 different camera models across many camera manufacturers. For some camera models, images from up to 5 instances are available in the dataset.

To generate the bona fide and morphed images, we first selected 315 images from the Dresden image database [98], consisting of 5 random images for every one of the 63 cameras. For the generation of the morphed image samples, we used the same morphing parameters as they would occur in a face morphing attack. In this experiment, they were obtained from applying the OpenCV with dlib approach on the FRGCv2 database, as described in Section V-A. With these parameters, we generated a total of 53, 362 morphed images from bona fide image pairs of different cameras. Finally, a patch of 320×320 pixels is cropped from the center of all bona fide and morphed images.

The results of applying the proposed PRNU-based face morphing system on these bona fide and morphed images are

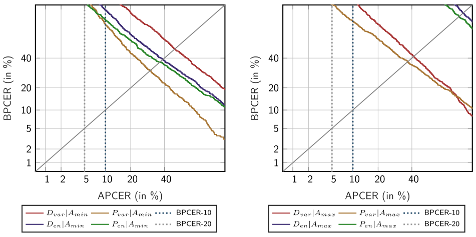
TABLE V
PERFORMANCE RESULTS IN TERMS OF D-EER (IN %) FOR DIFFERENT CONFIGURATIONS OF THE PROPOSED PRNU-BASED MORPHING ATTACK DETECTION SYSTEMS (CELL SIZE OF 10×10) AND 63 DIFFERENT CAMERAS FROM THE DRESDEN IMAGE DATABASE. "ALL" INDICATES THE RESULT FOR ALL CAMERA INSTANCES

Camera ID	D_{var}		P_{var}		D_{en}		P_{en}	
	A_{min}	A_{max}	A_{min}	A_{max}	A_{min}	A_{max}	A_{min}	A_{max}
Canon Ixus55 0	48.04	42.69	0.00	0.00	1.12	0.56	0.00	0.00
Canon Ixus70 0	30.00	15.61	0.00	0.00	6.94	14.82	0.00	0.00
Canon Ixus70 1	23.94	18.19	0.00	0.00	5.15	10.59	0.00	0.00
Canon Ixus70 2	20.32	13.58	0.00	0.00	11.28	13.97	0.00	0.00
Casio EXZ150 0	9.49	6.54	0.02	0.01	2.12	1.43	0.00	0.00
Casio EXZ150 1	16.33	10.58	0.05	0.01	1.27	2.35	0.00	0.00
Casio EXZ150 2	17.61	12.67	0.00	0.00	5.70	4.49	0.00	0.00
Casio EXZ150 3	14.96	9.04	0.02	0.00	2.55	2.35	0.00	0.00
Casio EXZ150 4	18.43	8.43	0.00	0.00	4.46	2.84	0.00	0.00
FujiFilm FinePixJ50 0	33.53	27.97	20.97	40.24	37.86	22.84	38.68	21.54
FujiFilm FinePixJ50 1	30.31	34.57	17.98	31.86	29.07	20.06	31.21	17.11
FujiFilm FinePixJ50 2	31.86	30.86	19.03	30.93	25.59	15.85	26.67	17.16
Nikon CoolPixS710 0	6.92	10.99	0.01	16.67	17.01	1.34	16.67	0.00
Nikon CoolPixS710 1	8.54	4.32	0.02	0.00	1.27	1.33	0.00	0.00
Nikon CoolPixS710 2	8.91	9.86	0.04	0.00	0.42	0.39	0.00	0.00
Nikon CoolPixS710 3	15.97	18.17	0.07	0.00	0.31	1.06	0.00	0.00
Nikon CoolPixS710 4	18.40	9.54	0.05	0.00	16.78	16.83	16.67	16.67
Nikon D200 0	22.62	25.77	1.91	0.21	0.41	3.53	1.06	7.03
Nikon D200 1	16.31	11.35	11.69	7.09	1.18	5.88	3.67	14.24
Nikon D70 0	43.25	44.70	0.66	1.35	0.96	2.04	1.57	1.92
Nikon D70 1	47.11	45.54	0.07	0.16	0.20	1.61	0.77	0.96
Nikon D70s 0	45.68	43.98	0.44	1.84	1.39	1.88	1.80	2.60
Nikon D70s 1	45.73	46.01	0.03	0.01	0.12	0.45	0.21	0.23
Olympus mju 1050SW 0	33.07	34.13	0.00	0.00	1.38	0.01	0.00	0.01
Olympus mju 1050SW 1	24.48	21.08	0.00	0.00	0.59	0.02	0.00	0.01
Olympus mju 1050SW 2	31.95	29.91	0.00	0.00	0.74	0.02	0.00	0.01
Olympus mju 1050SW 3	32.56	22.17	0.00	0.00	1.09	0.03	0.00	0.01
Olympus mju 1050SW 4	27.07	32.65	0.00	0.00	0.89	0.02	0.00	0.01
Panasonic DMCFZ50 0	23.07	16.60	19.24	37.37	29.59	20.76	34.51	19.29
Panasonic DMCFZ50 1	22.35	16.81	15.43	34.88	24.46	15.09	31.91	17.26
Panasonic DMCFZ50 2	18.50	17.86	19.06	34.75	32.00	19.37	33.98	19.26
Pentax OptioA40 0	38.86	38.45	0.87	12.58	17.15	1.53	14.19	2.80
Pentax OptioA40 1	44.65	41.29	2.36	11.13	8.40	8.60	8.28	6.74
Pentax OptioA40 2	47.85	41.52	0.20	2.91	4.73	0.81	2.78	0.70
Pentax OptioA40 3	47.75	44.23	0.00	0.01	0.04	1.05	0.00	0.01
Pentax OptioW60 0	38.69	27.98	0.00	0.00	34.42	20.41	0.00	0.00
Praktica DCZ59 0	1.02	0.24	40.17	48.61	13.32	19.56	33.23	34.97
Praktica DCZ59 1	0.27	0.56	40.27	49.07	16.38	19.07	34.91	34.34
Praktica DCZ59 2	0.89	0.57	41.56	49.43	15.49	16.54	36.56	32.43
Praktica DCZ59 3	0.64	0.30	41.19	48.01	12.17	19.87	31.85	35.41
Praktica DCZ59 4	1.19	0.67	40.87	48.01	14.85	18.78	35.05	34.87
Ricoh GX100 0	12.09	11.54	0.00	0.00	15.74	5.12	0.00	0.00
Ricoh GX100 1	16.12	13.53	0.00	0.00	9.28	1.96	0.00	0.00
Ricoh GX100 2	16.12	13.27	0.00	0.00	16.82	5.64	0.00	0.00
Ricoh GX100 3	13.34	12.14	0.00	0.00	14.83	5.62	0.00	0.00
Ricoh GX100 4	9.93	8.95	0.00	0.00	17.61	10.04	0.00	0.00
Rollei RCP7325XS 0	33.14	29.70	8.50	12.68	14.42	21.98	14.41	14.98
Rollei RCP7325XS 1	43.21	35.58	6.71	8.02	13.14	16.86	11.91	10.83
Rollei RCP7325XS 2	30.54	32.78	9.30	15.85	17.40	19.60	18.81	15.16
Samsung L74wide 0	4.71	3.58	3.53	2.83	0.00	0.05	0.48	1.03
Samsung L74wide 1	2.53	1.50	3.97	2.06	0.00	0.01	0.32	1.63
Samsung L74wide 2	4.35	2.00	3.55	3.05	0.00	0.02	0.62	1.40
Samsung NV15 0	26.59	13.21	0.35	0.67	0.00	0.58	0.13	0.01
Samsung NV15 1	15.59	14.08	0.23	0.11	0.97	0.82	0.07	0.00
Samsung NV15 2	21.48	11.80	0.29	0.40	0.08	0.37	0.00	0.01
Sony DSCH50 0	11.51	14.21	0.00	0.00	9.54	13.00	0.00	0.00
Sony DSCH50 1	8.30	5.31	0.00	0.01	1.09	4.57	0.00	0.00
Sony DSCT77 0	15.89	8.83	0.00	0.00	19.45	15.80	0.00	0.00
Sony DSCT77 1	9.79	11.90	0.00	0.00	26.03	17.94	0.00	0.00
Sony DSCT77 2	12.25	5.37	0.00	0.00	25.01	23.85	0.00	0.00
Sony DSCT77 3	32.13	24.66	0.00	2.01	1.56	6.83	0.00	0.00
Sony DSCW170 0	1.82	1.48	0.10	1.97	1.09	5.10	0.15	0.00
Sony DSCW170 1	3.17	1.49	0.07	4.19	0.12	3.05	0.06	0.00
All	**28.60**	**25.80**	**13.65**	**16.65**	**17.17**	**16.42**	**17.28**	**14.50**

presented in Table V. Looking at the overall results for all cameras at the bottom of the table, we obtain a D-EER of 13.65% with $P_{var}|A_{min}$ aggregation. For most cameras the detection error rate is very low. However, some cameras exhibit higher error rates of around 15−20% and cameras of a specific model (Practica DCZ59) even of up to 41.56%. We assume that this degradation might be caused by camera-specific non-unique artefacts, since the degradation mostly occurs for all cameras of the same model, as the mentioned Practica DCZ59 or FujiFilm FinePixJ50 and Panasonic DMCFZ50. Though, it has to be noted that the degradation does not persist among all investigated features, where a fusion of multiple features might yield improved performance and more consistent results. The other proposed features D_{en}, P_{var} and P_{en} also achieve

TABLE VI
PERFORMANCE RESULTS IN TERMS OF D-EER (IN %) FOR DIFFERENT CONFIGURATIONS OF THE PROPOSED PRNU-BASED MORPHING ATTACK DETECTION SYSTEMS (CELL SIZE OF 10×10) FOR THE PRINT-SCAN DATASET

Feature Extraction	Cell aggregation	D-EER
D_{var}	A_{min}	46.87
	A_{max}	41.07
P_{var}	A_{min}	**30.52**
	A_{max}	36.81
D_{en}	A_{min}	38.63
	A_{max}	49.97
P_{en}	A_{min}	36.51
	A_{max}	49.92



(a) A_{min}

(b) A_{max}

Fig. 11. DET curves for different configurations of the proposed morphing attack detection on the printed and scanned images for all morphing algorithms (OpenCV/dlib, FaceMorpher, FaceFusion, UBO).

respectable overall results between 14.5 and 28.6% D-EER. The histogram-based feature P_{var}, which is independent of any training data, show a better generalizability over the various cameras compared to the energy-based features D_{en} and P_{en}.

These results demonstrate that PRNU-based features in general are able to generalize well over a large number of different cameras and show promising results for a face morph detection scenario.

D. Printed and Scanned Images

In this last experiment, we look at the performance of the PRNU-based morph detection approach when applied to the Print-Scan dataset described in Section V-A. This scenario is very challenging for a PRNU-based approach, since the scanning process of the images embeds the scanner's PRNU within all scanned images, which might prevent the detection of the morphed images. The D-EER results are presented in Table VI.

We can observe, that the detection performance significantly drops for all proposed feature-aggregation combinations, where the best result is obtained with $P_{var}|A_{min}$ with a D-EER of 30.52%. Fig. 11 illustrates the DET plots for all proposed morph detection algorithms on the printed and scanned images, where all morphing algorithms, i.e., OpenCV/dlib, FaceMorpher, FaceFusion and UBO, have been included. These results show that the scanners PRNU leads to a detection performance degradation for the proposed PRNU-based approach, however $P_{var}|A_{min}$ is still able to discriminate bona fide and morphed images to some degree in this print and scan scenario.

VI. CONCLUSION

Face morphing attacks pose a serious security risk to face recognition systems. In this work, the potential PRNU analysis has been thoroughly analyzed for the challenging task of *no-reference* face morph detection. In comprehensive cross-database experiments for which different face morphing and image post-processing techniques have been applied, the proposed PRNU-based morphing attack detection system has been shown to outperform other state-of-the-art methods. Moreover, the feasibility of detecting morphed face images from printed and scanned image data has been investigated. Since the proposed system is based on a simple and minimal approach, further detection performance improvements can be expected by fusing multiple PRNU features and by a more sophisticated classification approach based on machine learning techniques.

In contrast to *differential* morphing attack detection schemes, e.g., [61], which additionally process a trusted live capture of a subject's face the proposed approach is particularly useful in cases where only a single potentially morphed face image is presented, e.g., digital transmission of a face image for issuance of an electronic travel document which turns out to be relevant in some countries. In other scenarios, e.g., facial recognition at ABC gates, the presented PRNU-based morphing attack detection scheme could be fused with other (differential) approaches to further improve the detection performance.

REFERENCES

[1] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Sep./Oct. 2014, pp. 1–7.

[2] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," *IEEE Access*, vol. 7, pp. 23012–23026, 2019.

[3] G. Wolberg, "Image morphing: A survey," *Vis. Comput.*, vol. 14, nos. 8–9, pp. 360–372, Dec. 1998.

[4] A. Patel and P. Lapsiwala, "Image morphing algorithm: A survey," *Int. J. Comput. Appl. (IJCA)*, vol. 5, no. 3, pp. 156–160, 2015.

[5] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch, "PRNU-based detection of morphed face images," in *Proc. IEEE 6th Int. Workshop Biometrics Forensics (IWBF)*, 2018, pp. 1–7.

[6] M. Ferrara, A. Franco, and D. Maltoni, "On the effects of image alterations on face recognition accuracy," in *Face Recognition Across the Imaging Spectrum*. Cham, Switzerland: Springer Int., 2016, pp. 195–222.

[7] D. J. Robertson, A. Mungall, D. G. Watson, K. A. Wade, S. J. Nightingale, and S. Butler, "Detecting morphed passport photos: A training and individual differences approach," *Cogn. Res. Princ. Implicat.*, vol. 3, no. 1, p. 27, Jun. 2018.

[8] U. Scherhag, C. Rathgeb, and C. Busch, "Towards detection of morphed face images in electronic travel documents," in *Proc. 13th IAPR Workshop Document Anal. Syst. (DAS)*, 2018, pp. 187–192.

[9] U. Scherhag, C. Rathgeb, and C. Busch, "Performance variation of morphed face image detection algorithms across different datasets," in *Proc. IEEE 6th Int. Workshop Biometr. Forensics (IWBF)*, Jun. 2018, pp. 1–6.

[10] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch, "PRNU variance analysis for morphed face image detection," in *Proc. 9th IEEE Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, 2018, pp. 1–9.

[11] R. Ramachandra, K. B. Raja, and C. Busch, "Detecting morphed face images," in *Proc. IEEE 8th Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, Sep. 2016, pp. 1–7.

[12] U. Scherhag, R. Ramachandra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "On the vulnerability of face recognition systems towards morphed face attacks," in *Proc. IEEE 5th Int. Workshop Biometr. Forensics (IWBF)*, Apr. 2017, pp. 1–6.

[13] R. Ramachandra, K. B. Raja, S. Venkatesh, and C. Busch, "Face morphing versus face averaging: Vulnerability and detection," in *Proc. IEEE Int. Joint Conf. Biometr. (IJCB)*, Oct. 2017, pp. 555–563.

[14] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 947–954.

[15] A. Agarwal, R. Singh, M. Vatsa, and A. Noore, "SWAPPED! Digital face presentation attack detection via weighted local magnitude pattern," in *Proc. IEEE Int. Joint Conf. Biometr. (IJCB)*, Oct. 2017, pp. 659–665.

[16] A. Asaad and S. Jassim, "Topological data analysis for image tampering detection," in *Digital Forensics and Watermarking*. Cham, Switzerland: Springer Int., 2017, pp. 136–146.

[17] S. Jassim and A. Asaad, "Automatic detection of image morphing by topology-based analysis," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 1007–1011.

[18] A. Makrushin, T. Neubert, and J. Dittmann, "Automatic generation and detection of visually faultless facial morphs," in *Proc. 12th Int. Joint Conf. Comput. Vis. Imag. Comput. Graph. Theory Appl.*, 2017, pp. 39–50.

[19] "Utrecht ECVP database," in *Proc. Eur. Conf. Vis. Perception*, Utrecht, The Netherlands, 2008.

[20] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann, "Modeling attacks on photo-ID documents and applying media forensics for the detection of facial morphing," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Security (IHMMSec)*, 2017, pp. 21–32.

[21] U. Scherhag, C. Rathgeb, and C. Busch, "Morph detection from single face image: A multi-algorithm fusion approach," in *Proc. Int. Conf. Biometr. Eng. Appl. (ICBEA)*, 2018, pp. 6–12.

[22] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, Apr. 1998.

[23] A. Martinez and R. Benavente, "The AR face database," Comput. Vis. Center, Universitat Autònoma de Barcelona, Barcelona, Spain, Rep. 24, Jun. 1998.

[24] L. Spreeuwers, M. Schils, and R. Veldhuis, "Towards robust evaluation of face morphing detection," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 1027–1031.

[25] N. Damer, A. M. Saladie, A. Braun, and A. Kuijper, "MorGAN: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network," in *Proc. 9th IEEE Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, 2018, pp. 1–10.

[26] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.

[27] L.-B. Zhang, F. Peng, and M. Long, "Face morphing detection using Fourier spectrum of sensor pattern noise," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.

[28] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, Jun. 2010.

[29] M. Hildebrandt, T. Neubert, A. Makrushin, and J. Dittmann, "Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps," in *Proc. IEEE 5th Int. Workshop Biometr. Forensics (IWBF)*, Apr. 2017, pp. 1–6.

[30] C. Seibold, A. Hilsmann, and P. Eisert, "Reflection analysis for face morphing attack detection," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 1022–1026.

[31] R. Ramachandra, S. Venkatesh, K. Raja, and C. Busch, "Detecting face morphing attacks with collaborative representation of steerable features," in *Proc. 3rd Comput. Vis. Image Process. (CVIP)*, 2018, pp. 1–11.

[32] R. Ramachandra, K. B. Raja, S. Venkatesh, and C. Busch, "Transferable deep-CNN features for detecting digital and print-scanned morphed face images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1822–1830.

[33] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Accurate and robust neural networks for security related applications exampled by face morphing attacks," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 1–16.

[34] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. IEEE 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, 2006, pp. 211–216.

[35] D. S. Ma, J. Correll, and B. Wittenbrink, "The chicago face database: A free stimulus set of faces and norming data," *Behav. Res. Methods*, vol. 47, no. 4, pp. 1122–1135, Jan. 2015.

[36] A. Kasinski, A. Florek, and A. Schmidt, "The PUT face database," in *Image Processing & Communications*. Bydgoszcz, Poland: UTP Univ. Sci. Technol., Jan. 2008.

[37] M. Grgic, K. Delac, and S. Grgic, "SCface—Surveillance cameras face database," *Multimedia Tools Appl.*, vol. 51, no. 3, pp. 863–879, Oct. 2009.

[38] L. Wandzik, G. Kaeding, and R. V. Garcia, "Morphing detection using a general- purpose face recognition system," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 1012–1016.

[39] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 807–813.

[40] M. Ngan, P. Grother, and K. Hanaoka, "Performance of automated facial morph detection and morph resistant face recognition algorithms," Nat. Inst. Stand. Technol., Gaithersburg, MD, USA, Rep., 2018.

[41] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[42] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 1363–1366.

[43] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2950–2959.

[44] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3025–3032.

[45] R. Ramachandra, S. Venkatesh, K. Raja, and C. Busch, "Towards making morphing attack detection robust using hybrid scale-space colour texture features," in *Proc. 5th Int. Conf. Identity Security Behav. Anal. (ISBA)*, 2019, pp. 22–24.

[46] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[47] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[48] C. Shu, X. Ding, and C. Fang, "Histogram of the oriented gradient for face recognition," *Tsinghua Sci. Technol.*, vol. 16, no. 2, pp. 216–224, Apr. 2011.

[49] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Rep. CMU-CS-16-118, 2016.

[50] T. Neubert, "Face morphing detection: An approach based on image degradation analysis," in *Digital Forensics and Watermarking*. Cham, Switzerland: Springer Int., 2017, pp. 93–106.

[51] A. Makrushin, C. Kraetzer, T. Neubert, and J. Dittmann, "Generalized Benford's law for blind detection of morphed face images," in *Proc. 6th ACM Workshop Inf. Hiding Multimedia Security (IH & MMSec)*, 2018, pp. 49–54.

[52] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2014, pp. 1–14.

[53] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[54] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Detection of face morphing attacks by deep learning," in *Digital Forensics and Watermarking*. Cham, Switzerland: Springer Int., 2017, pp. 107–120.

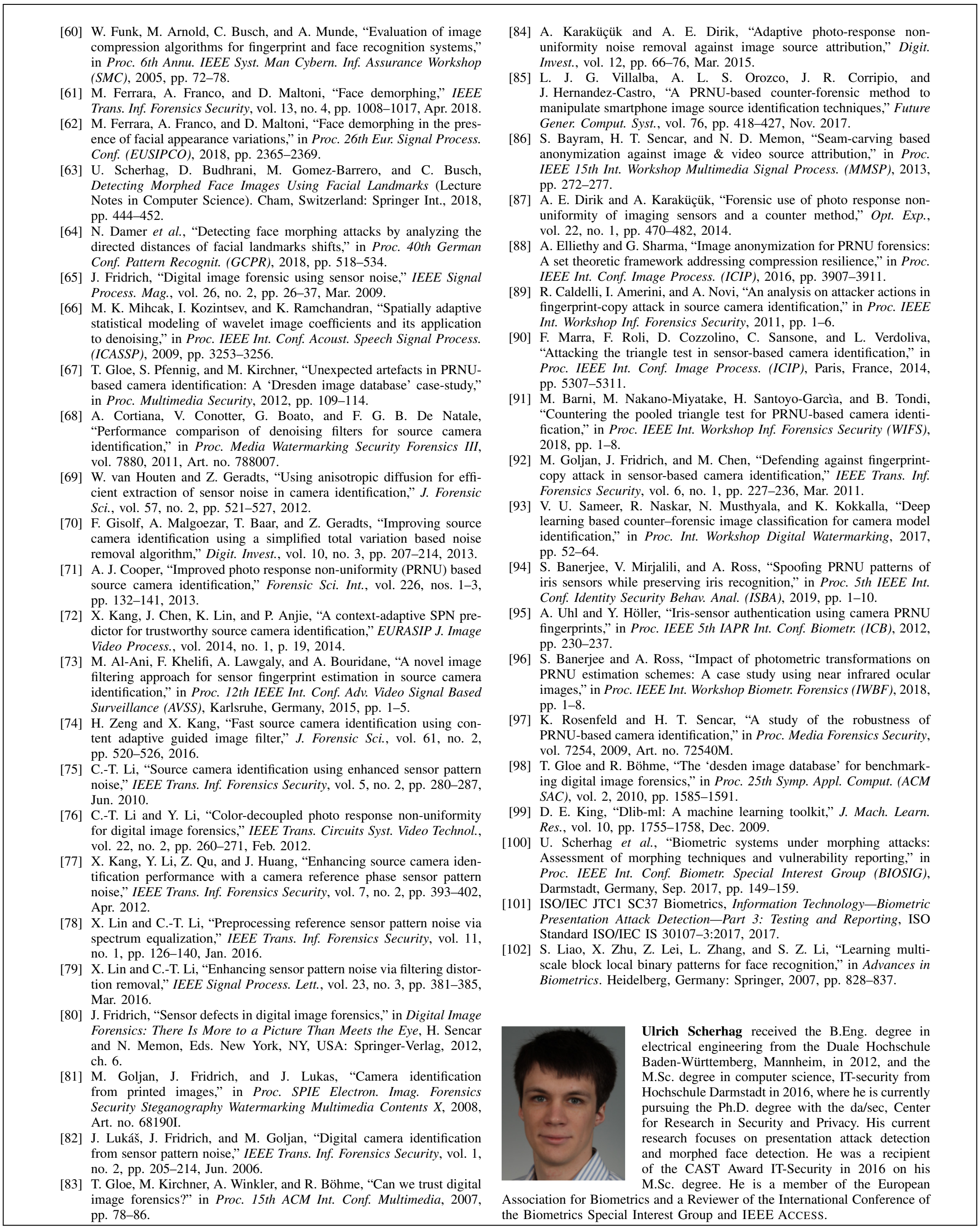
[55] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[56] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–12.

[57] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.

[58] *Information Technology—Biometric Data Interchange Formats—Part 5: Face Image Data*, Standard ISO/IEC JTC1 SC37 Biometrics, 2005.

[59] "Doc 9303, machine readable travel documents—Part 9: Deployment of biometric identification and electronic storage of data in MRTDs (7th edition)," ICAO, Montreal, QC, Canada, Rep. 7, 2015.



[60] W. Funk, M. Arnold, C. Busch, and A. Munde, "Evaluation of image compression algorithms for fingerprint and face recognition systems," in *Proc. 6th Annu. IEEE Syst. Man Cybern. Inf. Assurance Workshop (SMC)*, 2005, pp. 72–78.

[61] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1008–1017, Apr. 2018.

[62] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing in the presence of facial appearance variations," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, 2018, pp. 2365–2369.

[63] U. Scherhag, D. Budhrani, M. Gomez-Barrero, and C. Busch, *Detecting Morphed Face Images Using Facial Landmarks* (Lecture Notes in Computer Science). Cham, Switzerland: Springer Int., 2018, pp. 444–452.

[64] N. Damer *et al.*, "Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts," in *Proc. 40th German Conf. Pattern Recognit. (GCPR)*, 2018, pp. 518–534.

[65] J. Fridrich, "Digital image forensic using sensor noise," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 26–37, Mar. 2009.

[66] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2009, pp. 3253–3256.

[67] T. Gloe, S. Pfennig, and M. Kirchner, "Unexpected artefacts in PRNU-based camera identification: A 'Dresden image database' case-study," in *Proc. Multimedia Security*, 2012, pp. 109–114.

[68] A. Cortiana, V. Conotter, G. Boato, and F. G. B. De Natale, "Performance comparison of denoising filters for source camera identification," in *Proc. Media Watermarking Security Forensics III*, vol. 7880, 2011, Art. no. 788007.

[69] W. van Houten and Z. Geradts, "Using anisotropic diffusion for efficient extraction of sensor noise in camera identification," *J. Forensic Sci.*, vol. 57, no. 2, pp. 521–527, 2012.

[70] F. Gisolf, A. Malgoezar, T. Baar, and Z. Geradts, "Improving source camera identification using a simplified total variation based noise removal algorithm," *Digit. Invest.*, vol. 10, no. 3, pp. 207–214, 2013.

[71] A. J. Cooper, "Improved photo response non-uniformity (PRNU) based source camera identification," *Forensic Sci. Int.*, vol. 226, nos. 1–3, pp. 132–141, 2013.

[72] X. Kang, J. Chen, K. Lin, and P. Anjie, "A context-adaptive SPN predictor for trustworthy source camera identification," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, p. 19, 2014.

[73] M. Al-Ani, F. Khelifi, A. Lawgaly, and A. Bouridane, "A novel image filtering approach for sensor fingerprint estimation in source camera identification," in *Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, Karlsruhe, Germany, 2015, pp. 1–5.

[74] H. Zeng and X. Kang, "Fast source camera identification using content adaptive guided image filter," *J. Forensic Sci.*, vol. 61, no. 2, pp. 520–526, 2016.

[75] C.-T. Li, "Source camera identification using enhanced sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 280–287, Jun. 2010.

[76] C.-T. Li and Y. Li, "Color-decoupled photo response non-uniformity for digital image forensics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 2, pp. 260–271, Feb. 2012.

[77] X. Kang, Y. Li, Z. Qu, and J. Huang, "Enhancing source camera identification performance with a camera reference phase sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 393–402, Apr. 2012.

[78] X. Lin and C.-T. Li, "Preprocessing reference sensor pattern noise via spectrum equalization," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 126–140, Jan. 2016.

[79] X. Lin and C.-T. Li, "Enhancing sensor pattern noise via filtering distortion removal," *IEEE Signal Process. Lett.*, vol. 23, no. 3, pp. 381–385, Mar. 2016.

[80] J. Fridrich, "Sensor defects in digital image forensics," in *Digital Image Forensics: There Is More to a Picture Than Meets the Eye*, H. Sencar and N. Memon, Eds. New York, NY, USA: Springer-Verlag, 2012, ch. 6.

[81] M. Goljan, J. Fridrich, and J. Lukas, "Camera identification from printed images," in *Proc. SPIE Electron. Imag. Forensics Security Steganography Watermarking Multimedia Contents X*, 2008, Art. no. 68190I.

[82] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.

[83] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, "Can we trust digital image forensics?" in *Proc. 15th ACM Int. Conf. Multimedia*, 2007, pp. 78–86.

[84] A. Karaküçük and A. E. Dirik, "Adaptive photo-response non-uniformity noise removal against image source attribution," *Digit. Invest.*, vol. 12, pp. 66–76, Mar. 2015.

[85] L. J. G. Villalba, A. L. S. Orozco, J. R. Corripio, and J. Hernandez-Castro, "A PRNU-based counter-forensic method to manipulate smartphone image source identification techniques," *Future Gener. Comput. Syst.*, vol. 76, pp. 418–427, Nov. 2017.

[86] S. Bayram, H. T. Sencar, and N. D. Memon, "Seam-carving based anonymization against image & video source attribution," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, 2013, pp. 272–277.

[87] A. E. Dirik and A. Karaküçük, "Forensic use of photo response non-uniformity of imaging sensors and a counter method," *Opt. Exp.*, vol. 22, no. 1, pp. 470–482, 2014.

[88] A. Elliethy and G. Sharma, "Image anonymization for PRNU forensics: A set theoretic framework addressing compression resilience," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 3907–3911.

[89] R. Caldelli, I. Amerini, and A. Novi, "An analysis on attacker actions in fingerprint-copy attack in source camera identification," in *Proc. IEEE Int. Workshop Inf. Forensics Security*, 2011, pp. 1–6.

[90] F. Marra, F. Roli, D. Cozzolino, C. Sansone, and L. Verdoliva, "Attacking the triangle test in sensor-based camera identification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, 2014, pp. 5307–5311.

[91] M. Barni, M. Nakano-Miyatake, H. Santoyo-Garcìa, and B. Tondi, "Countering the pooled triangle test for PRNU-based camera identification," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, 2018, pp. 1–8.

[92] M. Goljan, J. Fridrich, and M. Chen, "Defending against fingerprint-copy attack in sensor-based camera identification," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 1, pp. 227–236, Mar. 2011.

[93] V. U. Sameer, R. Naskar, N. Musthyala, and K. Kokkalla, "Deep learning based counter–forensic image classification for camera model identification," in *Proc. Int. Workshop Digital Watermarking*, 2017, pp. 52–64.

[94] S. Banerjee, V. Mirjalili, and A. Ross, "Spoofing PRNU patterns of iris sensors while preserving iris recognition," in *Proc. 5th IEEE Int. Conf. Identity Security Behav. Anal. (ISBA)*, 2019, pp. 1–10.

[95] A. Uhl and Y. Höller, "Iris-sensor authentication using camera PRNU fingerprints," in *Proc. IEEE 5th IAPR Int. Conf. Biometr. (ICB)*, 2012, pp. 230–237.

[96] S. Banerjee and A. Ross, "Impact of photometric transformations on PRNU estimation schemes: A case study using near infrared ocular images," in *Proc. IEEE Int. Workshop Biometr. Forensics (IWBF)*, 2018, pp. 1–8.

[97] K. Rosenfeld and H. T. Sencar, "A study of the robustness of PRNU-based camera identification," in *Proc. Media Forensics Security*, vol. 7254, 2009, Art. no. 72540M.

[98] T. Gloe and R. Böhme, "The 'desden image database' for benchmarking digital image forensics," in *Proc. 25th Symp. Appl. Comput. (ACM SAC)*, vol. 2, 2010, pp. 1585–1591.

[99] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Dec. 2009.

[100] U. Scherhag *et al.*, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *Proc. IEEE Int. Conf. Biometr. Special Interest Group (BIOSIG)*, Darmstadt, Germany, Sep. 2017, pp. 149–159.

[101] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, ISO Standard ISO/IEC IS 30107–3:2017, 2017.

[102] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*. Heidelberg, Germany: Springer, 2007, pp. 828–837.

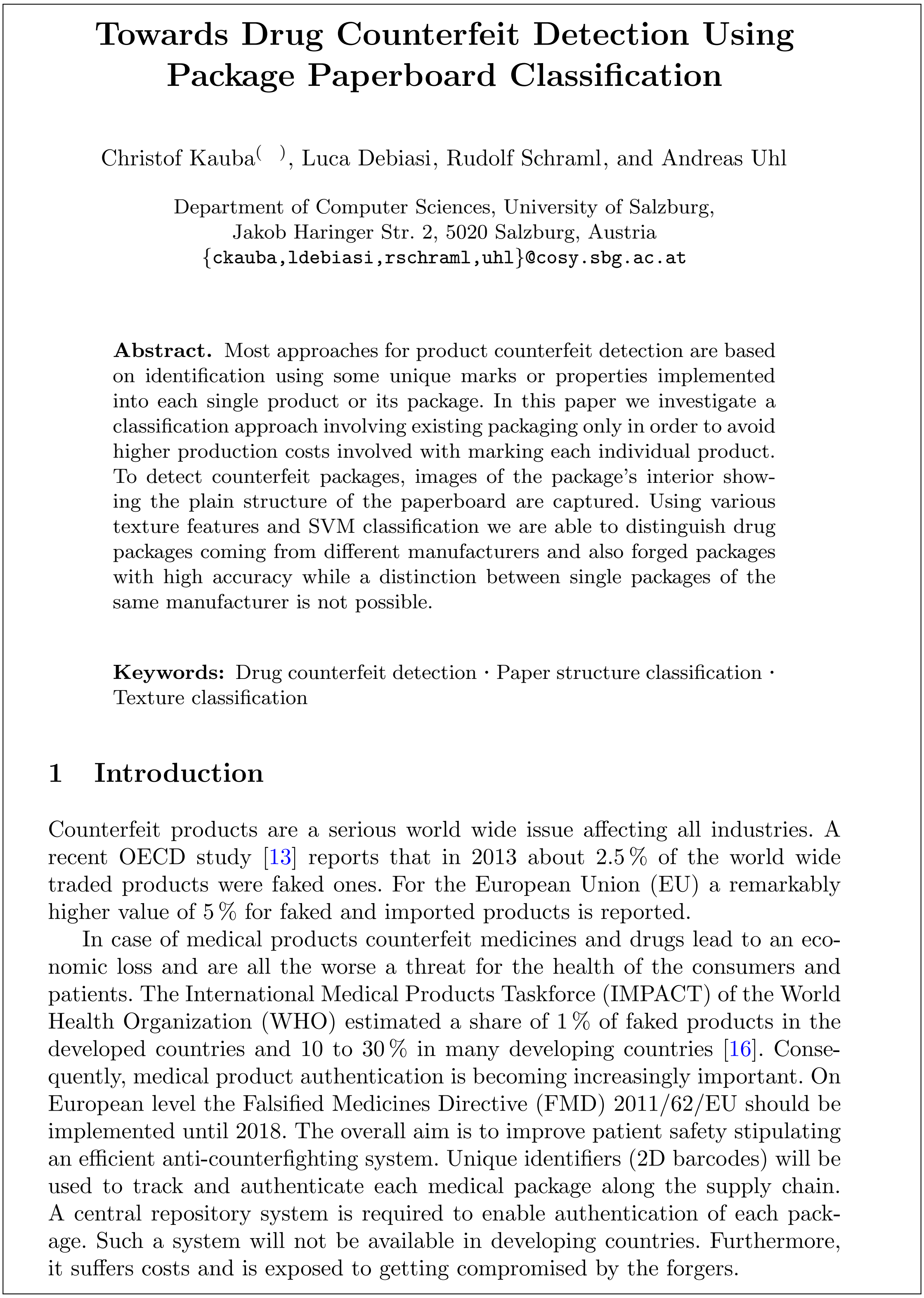
Ulrich Scherhag received the B.Eng. degree in electrical engineering from the Duale Hochschule Baden-Württemberg, Mannheim, in 2012, and the M.Sc. degree in computer science, IT-security from Hochschule Darmstadt in 2016, where he is currently pursuing the Ph.D. degree with the da/sec, Center for Research in Security and Privacy. His current research focuses on presentation attack detection and morphed face detection. He was a recipient of the CAST Award IT-Security in 2016 on his M.Sc. degree. He is a member of the European Association for Biometrics and a Reviewer of the International Conference of the Biometrics Special Interest Group and IEEE ACCESS.

Luca Debiasi received the B.Eng. and Dipl.-Ing. degrees in computer science from the Universität Salzburg in 2011 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the Multimedia Signal Processing and Security Lab, Department of Computer Sciences. His main research interests include digital image forensics, biometrics (hand- and finger-veins, iris, face), presentation attack detection, privacy enhancing technologies, and texture classification.

Christian Rathgeb is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt, Germany. He is a Principal Investigator with the Center for Research in Security and Privacy. He coauthored over 100 technical papers in the field of biometrics. His research includes pattern recognition, iris and face recognition, security aspects of biometric systems, secure process design, and privacy enhancing technologies for biometric systems. He was a recipient of the EAB—European Biometrics Research Award 2012, the Austrian Award of Excellence 2012, the Best Poster Paper Awards (IJCB'11, IJCB'14, and ICB'15), and the Best Paper Award Bronze (ICB'18). He is a member of the European Association for Biometrics (EAB), the Program Chair of the International Conference of the Biometrics Special Interest Group, and a editorial board member of *IET Biometrics*. He has served for various program committees and conferences (e.g., ICB, IJCB, BIOSIG, and IWBF) and journals as a reviewer (e.g., the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE, and *IET Biometrics*).

Christoph Busch is a Member of the Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway. He holds a joint appointment with the computer science faculty with Hochschule Darmstadt, Germany. He has been lecturing Biometric Systems with the Technical University of Denmark since 2007. He coauthored over 400 technical papers and has been a speaker at international conferences. He served for various program committees (NIST IBPC, ICB, ICHB, BSI-Congress, GI-Congress, DACH, WEDELMUSIC, and EUROGRAPHICS) and served for several conferences, journals, and magazines as reviewer (e.g., ACM-SIGGRAPH, the *ACM Transactions on Information and System Security*, the IEEE COMPUTER GRAPHICS AND APPLICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and the *Elsevier Journal Computers and Security*). He is also an Appointed Member of the editorial board of the *IET Biometrics* and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY journal. On behalf of Fraunhofer, he chairs the biometrics working group of the TeleTrusT association, as well as the German standardization body on Biometrics (DIN-NIA37). He is a Convenor of WG3 in ISO/IEC JTC1 SC37 on Biometrics and an Active Member of CEN TC 224 WG18.

Andreas Uhl received the Ph.D. degree from the Universität Salzburg, where he is currently a Professor with the Department of Computer Sciences. He has coauthored over 400 scientific publications. His research interests are in processing and analysis of visual data in general, and in biometric systems, multimedia security and forensics, and medical data analysis in particular. He acts as an Associate Editor of the *ACM Transactions on Multimedia Computing, Communications, and Applications*, *Signal Processing: Image Communication*, the *Journal of Visual Communication and Image Representation*, and *ETRI Journal*.



Towards Drug Counterfeit Detection Using Package Paperboard Classification

Christof Kauba( ), Luca Debiasi, Rudolf Schraml, and Andreas Uhl

Department of Computer Sciences, University of Salzburg,
Jakob Haringer Str. 2, 5020 Salzburg, Austria
{ckauba,ldebiasi,rschraml,uhl}@cosy.sbg.ac.at

Abstract. Most approaches for product counterfeit detection are based on identification using some unique marks or properties implemented into each single product or its package. In this paper we investigate a classification approach involving existing packaging only in order to avoid higher production costs involved with marking each individual product. To detect counterfeit packages, images of the package's interior showing the plain structure of the paperboard are captured. Using various texture features and SVM classification we are able to distinguish drug packages coming from different manufacturers and also forged packages with high accuracy while a distinction between single packages of the same manufacturer is not possible.

Keywords: Drug counterfeit detection · Paper structure classification · Texture classification

1 Introduction

Counterfeit products are a serious world wide issue affecting all industries. A recent OECD study [13] reports that in 2013 about 2.5 % of the world wide traded products were faked ones. For the European Union (EU) a remarkably higher value of 5 % for faked and imported products is reported.

In case of medical products counterfeit medicines and drugs lead to an economic loss and are all the worse a threat for the health of the consumers and patients. The International Medical Products Taskforce (IMPACT) of the World Health Organization (WHO) estimated a share of 1 % of faked products in the developed countries and 10 to 30 % in many developing countries [16]. Consequently, medical product authentication is becoming increasingly important. On European level the Falsified Medicines Directive (FMD) 2011/62/EU should be implemented until 2018. The overall aim is to improve patient safety stipulating an efficient anti-counterfighting system. Unique identifiers (2D barcodes) will be used to track and authenticate each medical package along the supply chain. A central repository system is required to enable authentication of each package. Such a system will not be available in developing countries. Furthermore, it suffers costs and is exposed to getting compromised by the forgers.

Another approach to verify the originality of a product is to use intrinsic features visible on the packaging or the product itself. For this work we focus on authentication of a medical product using intrinsic features from the packaging surface. Literature in this field relates to package fingerprinting based on the theory of physically uncloneable functions (PUFs). Paper PUFs use the fiber structure of paper as physical/intrinsic characteristic. The approaches presented in [1,3,10] show that the micro-structure in a certain region of a paper or package material is discriminative enough to identify it. Detailed investigations on paper identification, using a public available microstructure dataset [18], are presented in [4,5]. In [5] the authors explore the applicability of two approaches to overcome geometric distortions. The same approaches and a hybrid one are used to investigate package identification using mobile phones in [4]. Furthermore, in [6] a new feature descriptor for micro-structure identification using mobile phones is introduced. By comparing the performances for different PUFs the results in [20] indicate that the approach by [3] outperforms the approaches by [4,5,18] but it requires a commodity scanner. Thus, in [19] the authors showed that mobile devices and the camera built-in flash lights can also be used to capture images as required for [3].

As shown, research exclusively deals with identification of paper or packages. To the best of our knowledge no works which consider paper or package classification have been presented so far. Like in the work of [17] we assume that the fibre structure pattern of the packaging material is suited for classification, i.e. for a certain medical product the packaging fibre structure shows constant features. If so, one step for checking the authenticity of a medical product could be to assess if the packaging material is the same as used for the original product. To answer this question, we perform a preliminary study for nine different medical products from three different manufacturers and some forged packages for one medical product. The results of this work enable to draw conclusions which are a first step towards medical product authentication using the packaging material.

Section 2 introduces the basic concept of paper classification. The experimental setup and the data set acquisition are described in Sect. 3. Our experimental results together with a discussion of these results can be found in Sect. 4. Section 5 concludes this paper.

2 Paper Texture Classification

This section describes our proposed approach using paper texture classification for package counterfeit detection. The general procedure is the following: At first an image of the interior of the package is taken and several patches are extracted from random positions in the image. These patches are then preprocessed. Afterwards different features are extracted from the preprocessed patches. Based on these features a classifier returns a decision predicting the class a questioned image is belonging to (by utilizing a pre-trained SVM). The steps are explained in the following.

2.1 Image Acquisition

Several images of the package's interior are captured at different positions. For the image acquisition a Canon 70D (100 mm lens and flash light), mounted on a tripod, was utilized. The flashlight was placed besides the package. The camera is set to the smallest possible distance from the package (about 30 cm) trying to capture as most as possible of the paper's fibre-structure. An image of the acquisition setup can be seen in Fig. 1 together with an acquired image from the interior of a sample package.

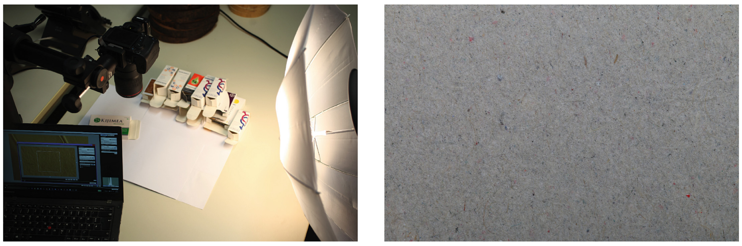


Fig. 1. Set-up for image acquisition of the fiber structure on the inside of a drug package (left) and acquired image sample (right).

2.2 Preprocessing

During preprocessing of the images a contrast limited adaptive histogram equalization (CLAHE) [21] is applied in order to improve contrast and enhance the paper structure. After this contrast enhancement all images are converted to grayscale and several patches are extracted from random positions in the images to reduce the computational effort and increase the amount of data that can be extracted from each package. Figure 2 shows the paper structure of different packages extracted from the random patches after preprocessing.

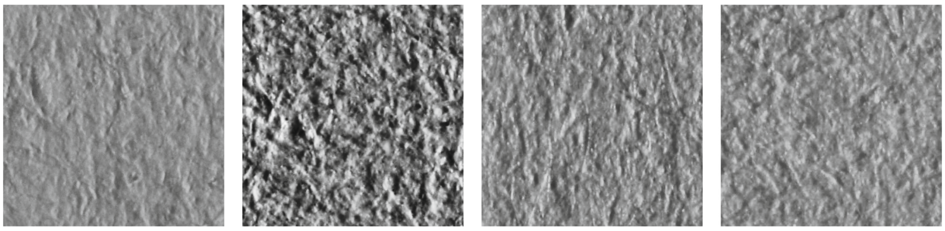


Fig. 2. Example preprocessed image patches

2.3 Feature Extraction Techniques

All techniques tested in this work are usually used for texture classification, image tampering detection and printer/paper identification and are applied on the preprocessed images taken from the inside of the package. The techniques utilized in this work are briefly described in the following list, further information on the single techniques can be found in the corresponding papers.

- Histogram
  Gray-level histogram of all pixels as the extracted feature.
- LBP: Local Binary Patterns
  The local binary patterns (LBP) by Ojala *et al.* [14] observe the variations of pixels in a local neighbourhood and are represented in a histogram.
- DMD: Dense Micro-block Difference
  Texture classification approach by Metha *et al.* [9] which captures the local structure from the image patches at high scales, but instead of the pixels small blocks which capture the micro-structure of the image are processed.
- RI-LPQ: Rotation-Invariant Local Phase Quantization
  The rotation-invariant local phase quantization (RI-LPQ) by Ojansivu *et al.* [15] consists of two stages: Estimation of the local characteristic orientation for a given image patch and directed descriptor extraction.
- Dense SIFT: Dense Scale Invariant Feature Transform
  Lowe [8] proposed a technique used in object recognition which is commonly known as scale invariant feature transform (SIFT). This technique is invariant to image scale and rotation and robust against various affine distortions, addition of noise, illumination changes and changes of the viewpoint.
- GLCM: Gray-level Co-occurence Matrix
  Mikkilineni *et al.* proposed to use gray-level co-occurence features for printer identification in [11]. The features model the spatial relationships among the pixels of an image to represent its texture information.
- WP: Weber Pattern
  In [12] Muhammad proposed a multi-scale local texture descriptor which was applied as part of an image forgery framework.
- BSIF: Binarized Statistical Image Features
  The Binarized Statistical Image Features (BSIF) proposed by Kannala *et al.* in [7] rely on pre-computed local image descriptors which efficiently encode texture information.
- LSB+JD: Least Significant Bitplane + Jaccard Distance
  Extraction of the images least significant bitplane (LSB-plane) and calculate the Jaccard distance between the LSB-planes of two images.

2.4 Classification Approach

The features extracted with the techniques described in the previous section are used to classify the images of the various kinds of drug packages.

The classifier is designed according to the improved Fisher vector (IFV) SVM classifier in [2]. The features are soft-quantized using a Gaussian mixture model (GMM), decorrelated and dimensionality reduced by PCA to obtain a Fisher vector (FV) encoding. A pre-trained linear SVM is then used to classify the IFV encoded features. The SVM is trained using a subset of the package's images which is subsequently not used for the testing (evaluation) step.

3 Experimental Settings

The following section describes the dataset used in this work, which contains images showing the paper structure of different forged and original drug packages. Furthermore a description of the two different dataset splits and our evaluation methodology to avoid overlapping between training and testing data is given.

3.1 Dataset

Unfortunately, only a limited number of drug packages was available for our work. In particular we have packages of 9 different kinds of drugs from 3 different manufacturers denoted by A, B and C.

For all 9 kinds of drugs we have genuine packages and for 2 of them we also have forged packages. The forged packages for the *Levitra* drug (ID 1) are real counterfeits confiscated by customs, while the forged packages for the *Neradin* drug (ID 8) have been purpose-made by the manufacturer of the drug.

Table 1 lists the number of genuine and forged packages for each kind of drug (ID 1...9). We acquired 10 to 20 slightly shifted and overlapping images from each of the packages' interiors from which 5 patches of 512×512 pixels are extracted at random position within each image. The extracted patches correspond to a section of approximately $4.1 \times 4.1\,\mathrm{mm}$, or $16.81\,\mathrm{mm}^2$, of the package. From this data we generated two distinct data sets to analyze two different issues using the paper structure of the packages:

1. Is it possible to distinguish different packages of the same manufacturer?
2. Is it possible to distinguish packages of different manufacturers?

The first data set, *SMDP* (Same Manufacturer Different Packages), contains images from packages of the same manufacturer, which correspond to the

Table 1. Number of genuine (G) and forged (F) packages in the data set with drug name, corresponding ID and manufacturer (MF).

ID	Name	# G	# F	MF
1	Levitra	3	4	A
2	Kijimea Reizdarm	2	0	B
3	Kijimea Immun	1	0	B
4	Kijimea Derma	2	0	B
5	Narumed	3	0	B
6	Deseo	4	0	B
7	Signasol	2	0	B
8	Neradin	4	2	B
9	Unistop	2	0	C

manufacturer B in Table 1. We only considered packages of this manufacturer since it is the only one from which we had more than one different type of drug package.

The second data set, *FGDM* (Forged and Genuine Different Manufacturers), contains images from all the packages, genuine and forged, from all manufacturers in Table 1.

3.2 Evaluation Methodology

To investigate the two questions of Sect. 3.1, we split the evaluation according to the two data sets SMDP and FGDM.

For the SMDP data set, where we want to find out if it is possible to distinguish between different types of drug packages from the same manufacturer, images having the same drug ID are defined as corresponding to the same class. A class thus can contain images from different packages of the same drug. Forged and genuine packages are furthermore split into different classes. This yields 8 different classes, because we have 7 different types of drug packages for manufacturer B and for one drug we also have 2 packages, which have been forged by the manufacturer.

To find out if it is possible to distinguish packages of different manufacturers (FGDM data set), images having the same manufacturer ID are defined as corresponding to the same class. Forged and genuine packages are again split into different classes for the *Levitra* drug produced by manufacturer A, but not for the *Neradin* drug of manufactuer B because these forgeries have been produced by the manufacturer and use the same material as the genuine packages. The different classes for the SMDP and FGDM data set are summarized in Table 2.

Table 2. Evaluation classes and corresponding IDs with number of packages

Name	# Packages	SMDP Class ID	FGDM Class ID
Levitra forged	4	-	1
Levitra genuine	3	-	2
Kijimea Reizdarm genuine	2	1	3
Kijimea Immun genuine	1	2	3
Kijimea Derma genuine	2	3	3
Narumed genuine	3	4	3
Deseo genuine	4	5	3
Signasol genuine	2	6	3
Neradin forged	2	7	3
Neradin genuine	4	8	3
Unistop genuine	2	-	4

The acquired images of the drug packages are slightly overlapping, this might lead to patches of the same image belonging to both, the training and the testing subset. Hence we used leave one package out (LOPO) for the selection of the training and testing images/patches: Training is done with randomly selected patches from all images except the images from one specific package. The patches for the testing subset are then randomly selected only out of images from this package. If there is only a single package in a class, like for the class with ID 2 in the case of the SMDP data set, the patches for this class are only used to train the classifier. Thus, no intra-class comparisons for this class exist and the average precision is not calculated and shown as 0 in the plots. By using the LOPO approach for the evaluation, the slight overlap of images from the same package does not introduce any bias to the results.

4 Experimental Results

This section presents the results of the conducted experiments and the conclusions made from those. We analysed the two cases, at first the separation according to manufacturer (FGDM) and second the separation of packages all from the same manufacturer (SMDP).

Table 3 lists the mean accuracies (mAcc) and mean average precisions (mAP) for both cases. The mean accuracy corresponds to the mean of the values of the confusion matrix diagonal. It can be seen that for FGDM the results for DenseSIFT and DMD are close to 100 % meaning that almost a perfect classification of the paper and thus the manufacturer is possible. Consequently, the true forgeries (corresponding to class 1) can be seperated from the other classes well.

Some example confusion matrices using a heat map for selected feature types (DMD, DenseSIFT and GLCM) can be seen in Figs. 3 and 5 for the FGDM and SMDP case, respectively. The numbers on the axes denote the classes according

Table 3. Mean accuracies (mAcc) and mean average precisions (mAP)

Data set	FGDM		SMDP	
Method	mAcc	mAP	mAcc	mAP
BSIF	0.428	0.403	0.138	0.171
DMD	0.97	1	0.328	0.423
DenseSIFT	0.91	1	0.37	0.476
GLCM	0.953	0.964	0.14	0.18
Histogram	0.603	0.662	0.145	0.176
LBP	0.758	0.863	0.265	0.272
LSB	0.71	0.818	0.113	0.182
RI-LPQ	0.842	0.888	0.158	0.226
WP	0.861	0.896	0.158	0.197

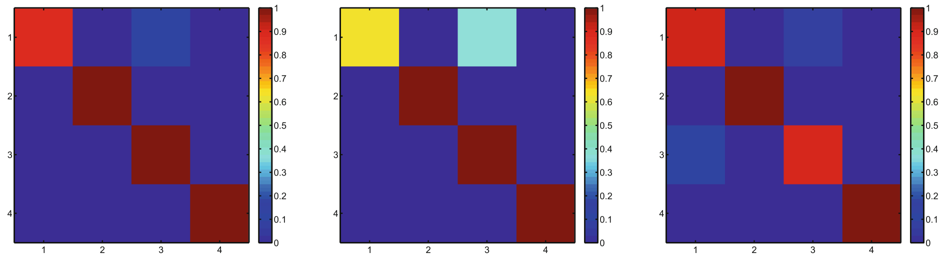


Fig. 3. Confusion matrix for DMD, DenseSIFT and GLCM in the FGDM case

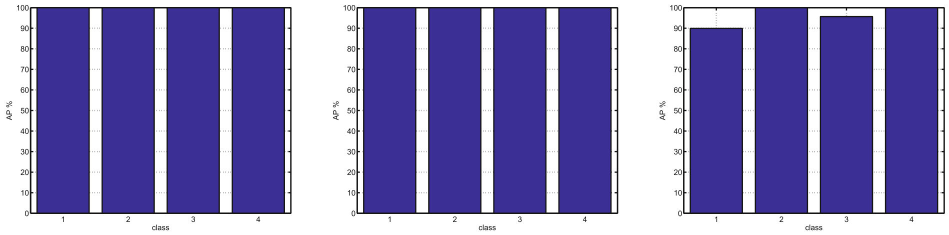


Fig. 4. Average precision for DMD, DenseSIFT and GLCM in the FGDM case

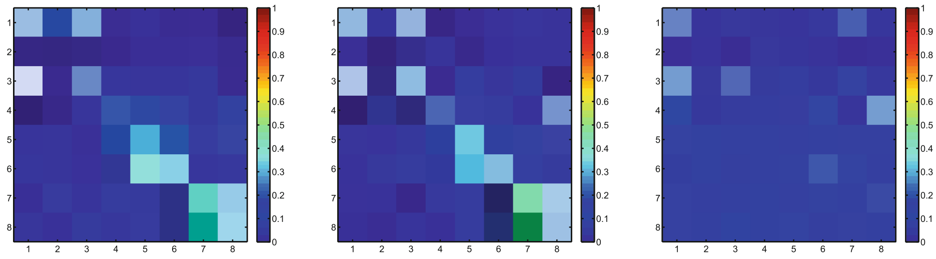


Fig. 5. Confusion matrix for DMD, DenseSIFT and GLCM in the SMDP case

to Table 2, which shows the correspondence of the class labels to the drug packages. Figures 4 and 6 show the corresponding average precision plots for FGDM and SMDP, respectively. These confirm that the recognition works well if the split is done according to different manufacturers and does not work if the split is done according to different drugs all from the same manufacturer.

We do not have any information about which kind of paper is used for the different drug packages. But the experimental results suggest (distinction between different types of drugs from the same manufacturer was not possible) that one manufacturer uses the same kind of paper and the same printing facility/printing process for his drug packages. As long as the forgers do not have access to the same kind of printing facility the genuine manufacturers utilizes, drug counterfeit detection is feasible using our proposed approach.

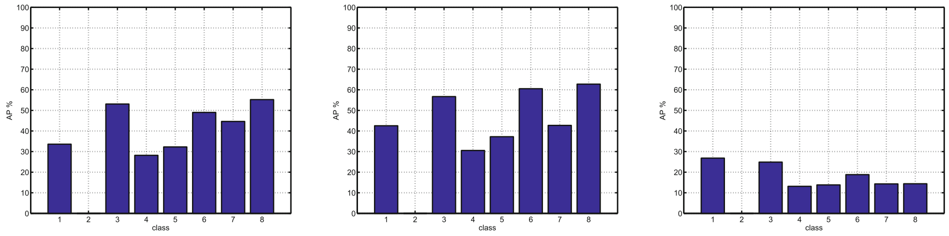


Fig. 6. Average precision for DMD, DenseSIFT and GLCM in the SMDP case

5 Conclusion

In this paper we investigated whether counterfeit drug package detection using texture classification based on the intrinsic paper texture is possible. The available data was split to investigate two different issues.

In the SMDP case (same manufacturer) a distinction between single packages of the same manufacturer was not possible. We concluded that this is not possible because all packages have very likely been produced using the same manufacturing process and therefore share a very similar paper structure.

In the FGDM case (different manufacturers) it was indeed possible to classify different genuine and forged packages with high accuracy. This indicates that it is possible to identify counterfeit packages not produced by the original manufacturer, since they are most likely being produced in a different manufacturing facility and hence do not share a similar paper structure. The class containing the forged packages and the classes containing genuine packages could all be clearly separated in this case.

This promising results however have to be taken with a grain of salt because of the small data set size and the availability of only a few real counterfeit packages. Hence the first step of our future work is the acquisition of more test data, i.e. a higher number of distinct types of drug packages and even more important more counterfeit and genuine packages of the same type of drug. In addition we want to acquire further information about the printing and manufacturing process of the packages.

Acknowledgments. This work is supported by the Munich based software venture eMundo which receives funding from the Central Innovation Program for SMEs by Germany's Federal Ministry of Economics and Technology (project "FakeFinder" no. ZIM-EP150145).

References

1. Buchanon, J., Cowburn, R., Jausovec, A.-V., Petit, S.: Forgery: fingerprinting documents and packaging. Nature **436**(7050), 475 (2005)
2. Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A.: Describing textures in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)

3. Clarkson, W., Weyrich, T., Finkelstein, A., Heninger, N., Halderman, J.A., Felten, E.W.: Fingerprinting blank paper using commodity scanners. In: Proceedings of the 30th IEEE Symposium on Security and Privacy, pp. 301–314 (2009)
4. Diephuis, M., Voloshynovskiy, S., Holotyak, T., Stendardo, N., Keel, B.: A framework for fast and secure packaging identification on mobile phones. In: Proceedings of SPIE Photonics West, Electronic Imaging, Media Forensics and Security V, San Francisco, USA, 23 January, 2014
5. Diephuis, M., Voloshynovskiy, S., Beekhof, F.: Physical object identification based on famos microstructure fingerprinting: comparison of templates versus invariant features. In: 8th International Symposium on Image and Signal Processing and Analysis, Trieste, Italy, 4–6 September, 2013
6. Diephuis, M., Voloshynovskiy, S., Holotyak, T.: Sketchprint: physical object microstructure identification using mobile phones. In: European Signal Processing Conference (EUSIPCO), Nice, France, 31 August - 4, September 2015
7. Kannala, J., Rahtu, E.: BSIF: binarized statistical image features. In: Proceedings of the 21st International Conference on Pattern Recognition, ICPR, Tsukuba, Japan, November 11–15, pp. 1363–1366 (2012)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
9. Mehta, R., Egiazarian, K.: Texture classification using dense micro-block difference (DMD). In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9004, pp. 643–658. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16808-1_43
10. Metois, E., Yarin, P., Salzman, N., Smith, J.: Fiberfingerprint identification. In: Proceedings of the 3rd Workshop on Automatic Identification, New York City, USA (2002)
11. Mikkilineni, A.K., Chiang, P.-J., Ali, G.N., Chiu, G.T.C., Allebach, J.P., Delp, E.J.: Printer identification based on graylevel co-occurrence features for security and forensic applications, vol. 5681, pp. 430–440 (2005)
12. Muhammad, G.: Multi-scale local texture descriptor for image forgery detection. In: IEEE International Conference on Industrial Technology (ICIT), pp. 1146–1151 (2013)
13. OECD and EUIPO: Trade in counterfeit and pirated goods, p. 138. OECD Publishing (2016)
14. Ojala, T., Pietikainen, M., Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol. 1, pp. 582–585, October 1994
15. Ojansivu, V., Rahtu, E., Heikkila, J.: Rotation invariant local phase quantization for blur insensitive texture analysis. In: 2008 19th International Conference on Pattern Recognition, ICpPR 2008, pp. 1–4, December 2008
16. W.H. Organization. International medical products taskforce - brochure. http://apps.who.int/impact/FinalBrochureWHA2008a.pdf?ua=1. Accessed 29 Apr 2016
17. Varma, M., Zisserman, A.: A statistical approach to material classification using image patch exemplars. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 2032–2047 (2009)
18. Voloshynovskiy, S., Diephuis, M., Beekhof, F., Koval, O., Keel, B.: Towards reproducible results in authentication based on physical non-cloneable functions: the forensic authentication microstructure optical set (famos). In: Proceedings of IEEE International Workshop on Information Forensics and Security, Tenerife, Spain, 2–5 December 2012

On the feasibility of classification-based product package authentication

Rudolf Schraml, Luca Debiasi, Cristof Kauba, Andreas Uhl
University of Salzburg, Department of Computer Sciences, 5020 Salzburg
{rschraml, ldebiasi, ckauba, uhl}@cosy.sbg.ac.at

***Abstract*—Depending on the product category the authenticity of a consumer good concerns economic, social and/or environmental issues. Counterfeited drugs are a threat to patient safety and cause significant economic losses. Different from physical-marking based approaches this work investigates authentication of drugs based on intrinsic texture features of the packaging material. Therefore, it is assumed that the packaging material of a certain drug shows constant but discriminative textural features which enable authentication, i.e. to prove if the packaging material is genuine or not. This objective requires considering a binary classification problem with an open set of negative classes, i.e. unknown and unseen counterfeits. In order to investigate the feasibility a novel drug packaging texture databases was acquired. The experimental evaluation of two basic requirements in texture classification serves as an evidence on the basic feasibility.**

I. Introduction

Counterfeiting is an economic issue affecting all industries. The OECD [1] reports that in 2013 2.5% of the worldwide traded products were counterfeited ones. For the European Union (EU) a remarkably higher value of 5% for counterfeited and imported products is reported. In case of medicals, counterfeits cause an economic loss and are moreover a potential threat to the consumer and patient health. On the European level, the Falsified Medicines Directive (FMD) 2011/62/EU should be implemented until 2018. The overall aim is to improve patient safety stipulating an efficient anti-counterfeiting system. The actual solution is based on product serialization, i.e. each package is assigned a unique identifier (e.g. 2D barcode) which enables to track and identify each medical package along the supply chain. Hence a central database is required to enable authentication of each package. Such a system will not be available in developing countries. Furthermore, it suffers costs and is exposed to getting compromised by forgers. For example, packages will have to be equipped with safety features in order to avoid tampering. Summarizing, serialization-based product authentication requires to adapt the production, shows significant risks & costs and cannot be implemented in a set of countries.

For this reason, we move from serialization to classification. This means that a product is authenticated based on constant but discriminative intrinsic features of the product or packaging material. Therefore, we target at pill drugs which are packaged in blisters and housed in a cardboard packaging. In [2] we showed that 9 different drugs from 3 manufacturers and some forged ones can be classified based on their cardboard packaging material, in a closed-set multi-class scenario. Results were promising and showed a classification accuracy of 100% for all 8 drugs. However, the testset is fairly small and drug package material authentication is a simplistic two-class (binary classification) problem, i.e. a drug is classified as being genuine or not. Contrasting to the setup in [2], package authentication has to be considered as an open set binary classification problem. In the training stage, the authentication system can capture only a limited subspace of other (known) drugs and forged packagings. It is a basic requirement that the authentication system is able to reject unseen counterfeited packages not known or available at the time of training. For a drug packaging authentication system this requires that a specific drug is distinguished from other known and unknown forgeries and drugs which is referred to as open set recognition. The general open set recognition problem has recently been addressed in the works of [3], [4], [5] which are outlined in Section II. Furthermore, in [6], [7], [8] the authors investigate the performances of the invented open set classification approaches in different applications.

In this work, we investigate the feasibility of a classification-based drug authentication system based on images of the cardboard packaging and top & bottom blister surface textures. Within this work the cardboard packaging texture and the blister top & bottom textures are referred to as modalities. A substantial drug packaging texture database, consisting of images from 45 drugs (multiple instances, i.e. multiple packages in the range of 1 and 15 per drug are acquired). Due to security concerns, strategic purposes and legal issues (toll, pharma industry) no forged packages were available.

So far, packaging or paper authentication refers to identification or serialization of each instance. These are based on the concept of physically unclonable functions (PUFs) which rely on the mapping between a challenge and response function depending on the physical nature of the object. PUFs are unclonable and unpredictable and thus ideally suited to implement identification-based anti-counterfeiting approaches. These either rely on extrinsic or intrinsic PUFs, i.e. which are attached to the product or can be derived from a part of the product itself. The encrypted PUF signature can be attached to the product enabling off-line authentication. [9], [10], [11] showed that the microstructure in a certain region of a paper or package material is discriminative enough to identify it (Paper PUFs). Detailed investigations on paper identification, using a publicly available microstructure dataset [12], are presented

WIFS‘2017, December, 4-7, 2017, Rennes, France. 978-1-5090-6769-5/17/$31.00 ©2017 European Union.

in [13], [14]. In [13] publicly availablethe authors explore the applicability of two approaches to overcome geometric distortions. The same approaches and a hybrid one are used to investigate package identification using mobile phones in [14]. Furthermore, in [15] a new feature descriptor for microstructure identification using mobile phones is introduced. By comparing the performances for different PUFs the results in [16] indicate that the approach by [11] outperforms the approaches by [12], [13], [14] but it requires a commodity scanner. Thus, in [17] the authors showed that mobile devices and the camera built-in flashlights can also be used to capture images as required for [11].

As identified in previous literature the fibre structure of paper or packaging material is positional highly unique and enables to identify single instances. The move from identification to classification, as done in this work, raises two fundamental research questions:

Positional invariance: Paper PUFs rely on the local uniqueness of the paper fibre texture. Thus, for the paper or cardboard packaging fibre structure it is not clear if (i) the fibre structure shows constant features across different regions and (ii) if those features are discriminative enough to distinguish between different types of paper or cardboard packaging.

Instance generalisation: The second question is a specialisation of the first for which the positional invariance is considered across different instances (i.e. packages) of a modality. Instance generalisation is a pre-requirement for a real-world application. For paper and packaging material it is not clear how the texture and the computed features vary between different instances, i.e. if a classifier which is trained with features from one instance is able to authenticate unseen features from another package instance and to distinguish them from other types of paper or cardboard packaging.

In this work, positional invariance and instance generalisation of the corresponding textural features are investigated for all three modalities. By considering these pre-requirements for classification-based drug packaging authentication, this work enables to draw fundamental conclusions. Based on the new insights the feasibility of a novel serialization-less anti-counterfeiting approach can be considered.

Section II introduces into open set drug package authentication: (i) Section II-A describes a possible scheme for an package authentication system and (ii) in Section II-B the open set recognition problem is considered in more detail. Section III introduces the acquired database. The classification pipeline is outlined in Section IV. Experiments and results are presented in Section V and Section VI concludes this paper.

II. Open-set drug package authentication

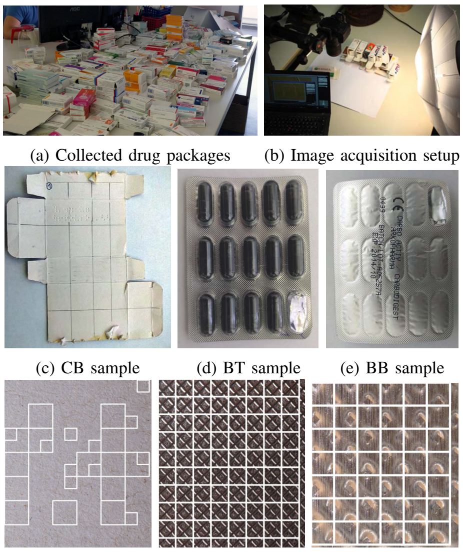
A. Drug package authentication system

For a given drug sample a mobile application guides the user to open/disassemble the drug packaging and to capture images of three different packaging modalities: The cardboard packaging texture I_{CB} as well as the textures visible on the top and bottom blister sides (I_{BT}, I_{BB}). Furthermore, the user is guided to capture the product code I_{PC} (e.g. EAN which is the European article number). All four images compose the authentication vector $\hat{AV} = (I_{CB}, I_{BT}, I_{BB}, I_{PC})$. I_{PC} is processed in order to determine the product code specifying the target product. I_{CB}, I_{BT}, I_{BB} are preprocessed (segmentation, image enhancement). For the resulting texture images T_{CB}, T_{BT}, T_{BB} a set of feature vectors $FV_{CB} = \{\hat{cb}_1, ..., \hat{cb}_i\}$, $FV_{BT} = \{\hat{bt}_1, ..., \hat{bt}_j\}$ and $FV_{BB} = \{\hat{bb}_1, ..., \hat{bb}_k\}$ are computed, where the number of feature vectors per modality i, j, k depends on the size of the preprocessed images and on the utilized feature extraction strategy. Based on the product code, the authentication system selects the corresponding precomputed classification models M_{CB}, M_{BT}, M_{BB} from a model repository. If the required models are not available on the device they could be requested from a remote repository. For each model M and a given feature vector $\hat{v}$ the prediction function $pF(M, v) = 1$ in case the vector is labelled as being genuine and -1 if not. For each model M_{CB}, M_{BT}, M_{BB} and the corresponding feature vector sets $FV_{CB}, FV_{BT}, FV_{BB}$ the prediction function is applied to all feature vectors which leads to the predictions for each modality of the packaging instance $P_{CB} = \{p_1, ..., p_i\}$, $P_{BT} = \{p_1, ..., p_j\}$ and $P_{BB} = \{p_1, ..., p_k\}$. Finally, a decision function $f(P_{CB}, P_{BT}, P_{BB}) = (v, p)$ needs to be defined, where $v \in \{1, -1\}$ gives the final authenticity vote of the authentication system and $p \in [0, 1]$ specifies a probability score for the final vote which are then presented to the user.

Such an authentication system relies on the assumption that different modalities of the packaging material of all instances from the same product show constant but discriminative features which enable to detect and distinguish the product from a known and unknown set (=open set) of other as well as from counterfeited products. For training of a classifier, only a limited subset of other drugs and available counterfeits is utilized. As a precondition for authentication, the classifier must be able to reject unseen data. This is a typical binary classification problem, either a given sample is labelled as genuine or not. The undefined set of unknown other classes leads to an open set recognition problem. This differs from closed-set classification where only known classes are separated from each other. Substantial efforts in the field of open set recognition were made in [3], [4], [5]. In [3] the authors introduce and formalize the open set recognition problem. Furthermore, in [3], [4], [5] the authors propose different SVM extensions which specifically address the open set recognition problem. In order to investigate the two research questions and as a consequence to prove the principal feasibility of an authentication system we base our experiments on the formalization of the open set recognition problem provided in [3], [4].

B. Formalization of the open set recognition problem

In [3] the authors define the *Open Space Risk* as $R_O(f) = \frac{\int_O f(x)dx}{\int_{S_O} f(x)dx}$. S_O needs to be considered as a large ball which is a subspace of the open space including all training samples. O is the open space. $f(x)$ is a recognition function where $f(x) = 1$ if x is recognized as the class of interest y and $f(x) = 0$ if



(a) Collected drug packages (b) Image acquisition setup

(c) CB sample (d) BT sample (e) BB sample

(f) CB image (g) BT image (h) BB image

Fig. 1: Image Acquisition Overview: 3rd Row: exemplary images showing the selected 128×128 and 256×256 image patches.

not. Consequently, the open set risk R_O is the fraction of the positively labelled open space in S_O compared to the positive labelled samples in O. The goal in open set recognition is to minimize the open space risk R_O whilst balancing it against the empirical (known) risk R_E computed over the available training data. Therefore, $\hat{P} = \{p_1, ..., p_n\}$ are samples from the positive training class P and $\hat{N} = \{n_1, ..., n_m\}$ are samples from a set of other known classes N. $\hat{N}$ is defined as the negative training data. U is the larger universe of negative unknown classes only utilized for evaluation and $E = \{e_1, ..., e_z\}, e_i \in P \cup N \cup U$ specifies all evaluation data. For a given training data $\hat{P} \cup \hat{N}$ and the open space and empirical risk functions R_O, R_E the open set recognition problem is to find a function f, where $f(x) > 0$ for positive recognitions, which minimizes the open set risk:

$$\underset{f}{arg\,min}\{R_O(f) + \lambda_r R_E(f(\hat{P} \cup \hat{N}))\} \tag{1}$$

where λ_r is a regularization constant.

Hence, open set recognition is a minimization problem which combines the open set and the empirical risk over the space of allowable recognition functions. Further, the empirical risk (i.e. the training error) can be optimized using predefined constraints. The stated minimization problem requires a set of known classes which are utilized for training and a set of known unknown classes in U which are only used for evaluation.

III. Drug Packagings Texture Database (DPT-DB)

For image acquisition, a large variety of drug packages were collected from different pharmacies (1st row in Fig. 1). From each drug package (=instance) the CB fibre texture on the inner raw side of the packaging, the BT texture (blister top side) and the BB texture (blister bottom side) were captured. For image acquisition a Canon 70D (100mm lens and flashlight), mounted on a tripod, was utilized (see Fig. 1b). From each CB,BT&BB instance images from different and non-overlapping sections were captured (e.g. Fig. 1c)). In total images for 45 drugs from 28 different producers were taken. For each drug between 1 and 15 package instances are available. All captured images were manually cropped ensuring that just texture remains.

IV. Classification Pipeline

Two different classification scenarios are considered: (i) CLASS to investigate the positional invariance of the CB,BT&BB texture. (ii) PACKAGE to prove instance invariance which is a step towards a real-world setup. In order to train and evaluate SVM-based classifiers data needs to be sampled and then partitioned into training (T) and evaluation (E) data. The amount of data (k) to be sampled is predefined for both scenarios. In this work, data relates to image texture patches of CB,BT&BB. For patch sampling, each CB,BT&BB image is subdivided into a grid which is specified by the size of the feature descriptor (e.g. 128×128 or 256×256 pixels). The 3rd row in Fig. 1 depicts sample images for CB,BT&BB for which the image patch grids are shown.

In case of CLASS k patches are sampled from all instances of each drug and modality. Contrary, for PACKAGE k patches are selected from each instance of each drug and modality. This is important in that for cross-validation the partitioning into T and E differs in principle as illustrated in Fig. 2. For CLASS the k patches of a drug and modality are partitioned so that different patches of each instance are included in T & E. On the other hand for PACKAGE the patches are partitioned instance-wise into T and E.

A. Feature vector computation

For each selected patch in CLASS or PACKAGE a set of discriminative features is computed. Prior to feature extraction Contrast Limited Adaptive Histogram Equalization (CLAHE) [18] is applied to each patch (parameters: block radius=50, bins=256, slope=40). Exemplary CLAHE enhanced patches are shown in the 1st and 2nd row of Fig. 3.

a) Feature Extraction: For the experiments feature extraction approaches producing low dimensional feature-vectors are utilized, mainly due to the fact that high dimensional features and feature encoding cause computational and memory issues when computing all classification configurations (CCs) for different SVMs, i.e. RAM & I/O limitations. We already did small-scale experiments on a subset of the CCs with SIFT,

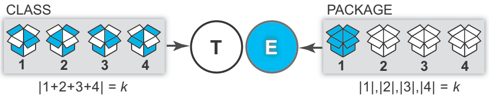


Fig. 2: Training (T) and evaluation (E) data sampling and partitioning strategies applied for CLASS and PACKAGE

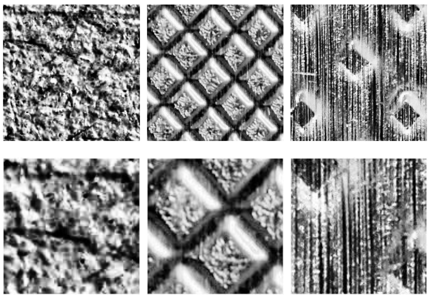


Fig. 3: Preprocessed patches of the CB,BT&BB images in Fig.1 - 1st Row: 256×256 pixels, 2nd Row: 128×128 pixels

SURF & feature encoding and the results indicate that the classification performance even increases.

The following features are utilized: Local Binary Pattern (LBP) [19], Local Ternary Pattern (LTP) [20], LiLBP (LiLBP) [21], Histogram of Gradients (HOG) [22], Dual Tree Complex Wavelet Transform (DTCWT) [23], Multifractal Spectrum (MFS) [24], Edge Co-Occurence Matrix (ECM) [25].

For each selected patch of CLASS & PACKAGE a feature vector for each listed feature extraction approach is computed.

B. *Classification Approaches*

For classification LIBSVM [26] and the open set extensions provided by [27] are utilized. From LIBSVM we use the ONE-CLASS and the C-SVC SVM (BINARY C-SVC)for one-class and binary classification, respectively. Additionally, as an approach specifically addressing open set recognition, the WSVM [4] is applied for binary classification. As the ONE-CLASS SVM uses a radial basis function (RBF) kernel, the same is chosen for BINARY C-SVC and WSVM.

In the experiments, the classification approaches are utilized to investigate a large set of different CCs. $D = \{d_1, ...d_{45}\}$ is the set of drugs and $DM = \{dm_1, ..., dm_{28}\}$ is the set of drug manufacturers in the testset where $fdm(d_i) : D \rightarrow DM$ specifies the drug manufacturer for each drug. $M = \{CB, BT, BB\}$ specifies the packaging modalities. $FE = \{fe_1, ..., fe_n\}$ is the set of feature extraction methods and $CS = \{CLASS, PACKAGE\}$ gives the classification scenarios. The feature vector sets for a certain drug $d \in D$ & modality $m \in M$, for the k-patches defined for the cla ssification scenario $cs \in CS$ computed with feature extraction method $fe \in F$, are given by $FV_{(d,m,f,cs)} = \{fv_1, ..., fv_k\}$. Following, a specific CC is defined by the tuple

$$CC = (d \in D, m \in M, fe \in FE, cs \in CS) \qquad (2)$$

where d specifies the target drug which should be authenticated. The respective set of feature vector sets for CC is given by $FV_{CC} = \{FV_{(d_1,m,f,cs)}, ..., FV_{(d_{45},m,f,cs)}\}$ which is composed by the CC specific feature vector sets from each drug. The positive training data $P_{CC} = FV_{(d,m,f,cs)}$ is specified by the target drug d in CC. The negative training data $N_{CC} = \{FV_{CC}\} \setminus \{FV_{(d,m,f,cs)}\}$ is composed by all feature vector sets of all other drugs. The positive and negative training data P_{CC}, N_{CC} are then used for nested cross-validation using a specific classification approach.

C. *Cross-fold validation*

Optimization is crucial as the standard LIBSVM parameters did not succeed in our experiments. Therefore, cross-validation (CV) strategies have been carefully designed and employed in order to optimize the SVM parameters and to strictly avoid that training data is used for evaluation.

Therefore, the negative training data is split into known negatives KN_{CC} and unknown negatives $UN_{CC} = N_{CC}/KN_{CC}$. Therefore, for KN_{CC} the feature vector sets from a fixed number of drugs (e.g. 6) are selected, where the manufacturers are different to the target drug manufacturer of d in CC. Now, a set of positive training data P_{CC}, a set of known negatives KN_{CC} and unknown negatives UN_{CC} is available. Based on P_{CC}, KN_{CC}, UN_{CC} nested CV procedures for CLASS and PACKAGE are defined as illustrated in Fig.4.

For CLASS, we apply a k-fold data split strategy, i.e. P_{CC}, KN_{CC} are class-wise split into k-folds $\{P_1, ..., P_k\}$ and $\{KN_1, ..., KN_k\}$, i.e. all drug classes are distributed equally in the k folds. In the outer loop, we iterate over the k positive and k negative known data folds. Thereby, the ith positive and jth negative was selected for evaluation. The evaluation set is given by $E_{i,j} = P_i \cup KN_j \cup UN_{CC}$ and the training set by $Ti, j = \{P_1, ..., P_k\} \setminus \{P_i\} \cup \{KN_1, ..., KN_k\} \setminus \{KN_j\}$. Thus a large set of known unknown drugs UN_{CC} are used only for evaluation. Note that $|\{KN_1, ..., KN_k\} \setminus \{KN_j\}|$ is reduced to the same size of the positive training data $|P_i|$ in a classwise manner. For each $T_{i,j}$ in the inner CV loop the best hyperparameters are determined in a grid search. Same as in the outer loop, k-fold validation is performed repeatedly in order to test a set of SVM parameters. For the ONE-CLASS SVM just the positive samples in $T_{i,j}$ are split into k-folds and the known negative training samples are only used for validation. As a measure for the performance the F-Measure is utilized which is well suited to balance between specialisation and generalisation in binary classification tasks. For the binary SVM approaches, each prediction is assigned a probability. In the inner loop, the probabilities are used to determine a threshold which maximizes the F-Measure. The SVM parameters delivering the highest F-Measure (and the probability threshold in case of binary SVMs) are selected for

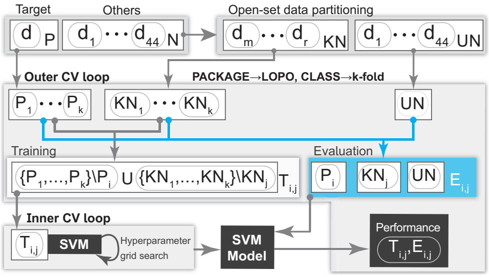


Fig. 4: Cross-validation scheme for CLASS and PACKAGE

CC	CLASS						PACKAGE					
	128×128			256×256			128×128			256×256		
	CB	BT	BB	CB	BT	BB	CB	BT	BB	CB	BT	BB
ONE-CLASS	LTP 0.83 ±7.9	LTP 0.9 ±6.2	LTP 0.92 ±5.8	LTP 0.91 ±4.4	LTP 0.85 ±13.6	LBP 0.87 ±13.5	LBP 0.81 ±8.7	LBP 0.86 ±6.3	LTP 0.84 ±11.3	LTP 0.85 ±9.1	LBP 0.88 ±5.0	LBP 0.85 ±7.1
BINARY	LTP 0.88 ±6.9	$LiLBP$ 0.94 ±3.2	LTP 0.93 ±4.1	LTP 0.91 ±5.2	$LiLBP$ 0.92 ±9.0	LTP 0.93 ±5.0	LTP 0.82 ±9.5	LTP 0.92 ±3.7	LTP 0.87 ±8.9	LTP 0.85 ±5.5	LTP 0.94 ±5.7	$LiLBP$ 0.87 ±10.0
WSVM	LTP 0.86 ±7.6	LTP 0.93 ±4.1	LTP 0.93 ±4.3	$LiLBP$ 0.88 ±6.0	LTP 0.88 ±7.6	MFS 0.88 ±9.1	LTP 0.85 ±8.2	LTP 0.91 ±4.2	$LiLBP$ 0.85 ±9.2	$LiLBP$ 0.83 ±8.5	LTP 0.89 ±8.7	$LiLBP$ 0.84 ±10.1

TABLE I: Classification performances: For each configuration the mean F-Measure (CLASS=45 & PACKAGE=8 target drugs) and the StDev for the best feature are presented. BEST CLASS/ PACKAGE configurations for each modality are layered green.

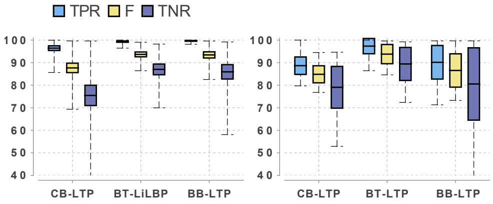
the outer loop. Finally, the SVM approach is trained with $T_{i,j}$ (for ONE-CLASS only the positive data P_i is utilized) and the selected hyper parameters from the inner CV loop. The trained model is evaluated using the evaluation data $E_{i,j}$ and probability threshold in case of binary SVMs.

For PACKAGE, a nested leave-one-package-out (LOPO) CV procedure is applied. Thereby, P_{CC} is split into k-folds in a package-wise manner, where k is given by the number of packages in P_{CC}, i.e. the number of available packages from the target drug. KN_{CC} is reduced to contain a fixed number of feature vectors from each class which are sampled package-wise. Furthermore, for KN_{CC} the features of each drug are split into two folds KN_1, KN_2 package-wise. Same as for CLASS, in the outer CV loop we iterate over the i positive and the $j = 2$ known negative training folds Ti, j and evaluate it with $E_{i,j}$, as done in the CLASS scenario. For ONE-CLASS in the inner CV loop the same procedure as for the outer loop is applied. However, for binary SVMs the inner CV loop has been adopted to better match the open set recognition problem. Therefore, the known negative training data in Ti, j is split classwise into two folds TKN_1 and TKN_2. One fold simulates known negatives and the other one unknown negatives in the inner loop. While the known negatives are further used for training and validation, the unknown negatives are just used for validation. This strategy adapts the inner CV loop and the parameter grid search to the open set recognition problem and is supposed to minimize the difference between the inner CV validation- and the outer CV evaluation-error.

V. Experiments

A. Experimental setup

All classification approaches (Section IV-B) were utilized to cross-validate all CC combinations (Eq.2) using the CS-specific CV strategies (Section IV-C). For both, PACKAGE and CLASS a patch number k of 500 is set. For CLASS the outer and inner CV loops are iterated twice and the data is split into 2-folds. In case of PACKAGE LOPO is performed for all package instances of drugs with at least 5 instances. For each LOPO CV the positive data is split into 2-folds, in the inner and outer CV loop. For both CSs, 5 drugs are selected for the known negative training data KN_{CC}. In order to enable a fair evaluation, all data splits for CLASS & PACKAGE are stored and reused for different features and classification approaches.



(a) CLASS - 128×128 (b) PACKAGE - 256×256

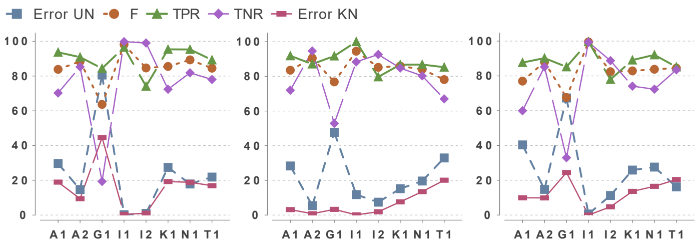
Fig. 5: CLASS vs. PACKAGE (Binary C-SVC): TPR = $\frac{TP}{TP+FN}$, TNR = $\frac{TN}{TN+FP}$ [Y-Axis: Mean accuracies, min, max and variance in %]

B. Results and discussion

Table I provides an overview of the results for each classification scenario, different patch sizes, modalities and SVMs. For CLASS the averaged results over all 45 drugs are shown. In case of PACKAGE, mean values for drugs with at least 5 instances are shown.

Considering positional invariance, the results for the best (green layered) CLASS configurations show high mean F-Measures over 0.9. This indicates that the textures from all three modalities show constant but highly discriminative features which enable to recognize the same drug class and to distinguish it from other classes. Regarding the question of instance invariance, the F-Measures for the best PACKAGE configurations provide an evidence on the feasibility of a drug package authentication system. The PACKAGE results show that the textural features are constant across different instances for all three modalities. This is a basic requirement for a classification-based authentication system. Although only low-level features have been utilized, the achieved F-Measures are very promising. Most of the best results for both scenarios and the different modalities were achieved with the BINARY C-SVC Y SVM. Fig. 5 provides a more detailed view on the BINARY C-SVC CLASS and PACKAGE results for the best features from each modality. Thereby, it is clearly visible that the performance decreases in case of the more difficult PACKAGE scenario. Furthermore, the comparison between the class accuracy (=true positive rate - TPR) and the others accuracy (=true negative rate - TNR) shows that for all results a higher class accuracy is achieved.

Finally, Fig. 6 shows accuracies and errors for PACKAGE,CB and all SVMs for the best features. For each tested drug (=8) and all SVMs, results show that the error for known data (KN=seen in training) is lower than the error for unknown data (UN=open set). Considering the different



(a) ONE-CLASS (b) BINARY C-SVC (c) WSVM

Fig. 6: PACKAGE (256×256) – SVM performance comparison for CB and all target drugs with more than 5 instances (=8 drugs): Accuracies (TPR,TNR) and recognition errors for the unseen data (Error UN) and seen training data (KN) are shown [X-Axis: Target drug (d) ids: e.g. A1 = manufacturer A+drug number].

SVMs, the accuracies and errors for ONE-CLASS and WSVM vary more compared to the per-drug results of the BINARY C-SVC. Furthermore, the WSVM does not outperform the classical BINARY C-SVC in terms of achieving a lower error for recognizing unknown data (UN).

VI. Conclusion

Results showed that textural features of drug packaging material are constant and highly discriminative. Very important, the experiments indicate that a classifier can be trained with a set of known instances and is able to authenticate unseen instances.

In future work, we will use high-level features, feature encoding and fusion techniques and it is planned to employ deep learning techniques. Furthermore, causes for classification errors need to be investigated in detail, e.g. in case of a high false positive rate it can be that other drugs from the same manufacturer have the same packaging material.

References

[1] OECD and EUIPO, “Trade in counterfeit and pirated goods,” *OECD Publishing*, p. 138, 2016.

[2] C. Kauba, L. Debiasi, R. Schraml, and A. Uhl, “Towards drug counterfeit detection using package paperboard classification,” in *Procs. of the 17th Pacific-Rim Conf. on Multimedia (PCM'16)*, vol. 9917. Springer LNCS, 2016, pp. 136–146.

[3] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult, “Towards open set recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 35, 2013.

[4] W. J. Scheirer, L. P. Jain, and T. E. Boult, “Probability models for open set recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 36, November 2014.

[5] L. P. Jain, W. J. Scheirer, and T. E. Boult, “Multi-class open set recognition using probability of inclusion,” in *The European Conf. on Computer Vision (ECCV'14)*, 2014.

[6] B. Heflin, W. J. Scheirer, and T. E. Boult, “Detecting and classifying scars, marks, and tattoos found in the wild,” in *The IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2012.

[7] F. O. Costa, E. Silva, M. Eckmann, W. J. Scheirer, and A. Rocha, “Open set source camera attribution and device linking,” *Pattern Recognition Letters*, vol. 36, April 2014.

[8] A. Rattani, W. J. Scheirer, and A. Ross, “Open set fingerprint spoof detection across novel fabrication materials,” *IEEE Trans. on Information Forensics and Security (T-IFS)*, vol. 10, 2015.

[9] E. Metois, P. Yarin, N. Salzman, and J. Smith, “Fiberfingerprint identification,” in *Procs. of the 3rd Workshop on Automatic Identification*, 2002.

[10] J. Buchanon, R. Cowburn, A.-V. Jausovec, S. Petit, D., G. P., Xiong, D. Atkinson, K. Fenton, D. Allwood, and T. Bryan, “Forgery: Fingerprinting documents and packaging,” *Nature*, vol. 436, p. 475, 2005.

[11] W. Clarkson, T. Weyrich, A. Finkelstein, N. Heninger, J. A. Halderman, and E. W. Felten, “Fingerprinting blank paper using commodity scanners,” in *30th IEEE Symp. on Security and Privacy*, 2009, pp. 301–314.

[12] S. Voloshynovskiy, M. Diephuis, F. Beekhof, O. Koval, and B. Keel, “Towards reproducible results in authentication based on physical non-cloneable functions: The forensic authentication microstructure optical set (famos),” in *Procs. of IEEE Int. Workshop on Information Forensics and Security (WIFS'12)*, 2012.

[13] M. Diephuis, S. Voloshynovskiy, and F. Beekhof, “Physical object identification based on FAMOS microstructure fingerprinting: comparison of templates versus invariant features,” in *8th Int. Symposium on Image and Signal Processing and Analysis*, Trieste, Italy, September, 4-6 2013.

[14] M. Diephuis, F. Beekhof, S. Voloshynovskiy, T. Holotyak, N. Standardo, and B. Keel, “A framework for fast and secure packaging identification on mobile phones,” in *Procs. of SPIE Photonics West, Electronic Imaging, Media Forensics and Security V*, 2014.

[15] M. Diephuis, S. Voloshynovskiy, and T. Holotyak, “Sketchprint: Physical object micro-structure identification using mobile phones,” in *European Signal Processing Conf. (EUSIPCO'15)*, 2015.

[16] C. W. Wong and M. Wu, “A study on PUF characteristics for counterfeit detection,” in *IEEE Int. Conf. on Image Processing (ICIP'15)*, 2015, pp. 1643–1647.

[17] C. Wong and M. Wu, “Counterfeit detection using paper PUF and mobile cameras,” in *IEEE Int. Workshop on Information Forensics and Security (WIFS'15)*, 2015, pp. 1–6.

[18] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics Gems IV*. Morgan Kaufmann, 1994, pp. 474–485.

[19] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution Gray-Scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[20] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” in *Analysis and Modelling of Faces and Gestures*, ser. LNCS, vol. 4778, 2007, pp. 168–182.

[21] Z. Li, G. Liu, Y. Yang, and J. You, “Scale- and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift,” *IEEE Trans. on Image Processing*, vol. 21, no. 4, pp. 2130–2140, 2012.

[22] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Procs. of the IEEE Conf. on Computer Vision and Pattern Recognition, (CVPR'05)*, vol. 1, 2005, pp. 886–893.

[23] S. Hatipoglu, S. Mitra, and N. G. Kingsbury, “Texture classification using dual-tree complex wavelet transform,” in *7th Int. Conf. on Image Processing and Its Applications*, vol. 1, Jul. 1999, pp. 344–347.

[24] Y. Xu, H. Ji, and C. Fermüller, “Viewpoint invariant texture description using fractal analysis,” *Int. Journal of Computer Vision*, vol. 83, no. 1, pp. 85–100, 2009.

[25] R. Rautkorpi and J. Iivarinen, “A novel shape feature for image classification and retrieval,” in *Procs. of the Int. Conf. on Image Analysis and Recognition (ICIAR'04)*, ser. LNCS, vol. 3211, 2004, pp. 753–760.

[26] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[27] Lalit Jain, “Openset libsvm extensions,” https://github.com/ljain2/libsvm-openset, 2014, [last accessed: 13.06.2017].

Real or Fake: Mobile Device Drug Packaging Authentication

Rudolf Schraml, Luca Debiasi, Andreas Uhl
University of Salzburg
Department of Computer Sciences
rschraml@cs.sbg.ac.at,ldebiasi@cs.sbg.ac.at,uhl@cs.sbg.ac.at

ABSTRACT

Shortly, within the member states of the European Union a serialization-based anti-counterfeiting system for pharmaceutical products will be introduced. This system requires a third party enabling to track serialized and enrolled instances of each product from the manufacturer to the consumer.

An alternative to serialization is authentication of a product by classifying it as being real or fake using intrinsic or extrinsic features of the product. Thereby, one approach is packaging material classification using images of the packaging textures. While the basic feasibility has been proven recently, it is not clear if such an authentication system works with images captured with mobile devices. Thus, in this work mobile device drug packaging authentication is investigated. The experimental evaluation provides results on single- and cross-sensor scenarios. Results indicate the principal feasibility and acknowledge open issues for a mobile device drug packaging authentication system.

ACM Reference Format:
Rudolf Schraml, Luca Debiasi, Andreas Uhl. 2018. Real or Fake: Mobile Device Drug Packaging Authentication. In *IH&MMSec '18: 6th ACM Workshop on Information Hiding and Multimedia Security, June 20–22, 2018, Innsbruck, Austria.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3206004.3206016

1 INTRODUCTION

As the global markets get flooded with counterfeited products regulations and technical solutions for product authentication get implemented in various sectors of the economy. According to a report by the European Intellectual Property Office 4.4% of the sales and € 10 billion in the pharmaceutical sector correspond to counterfeited medicines [2]. Moreover, counterfeit drugs pose a significant risk to consumer or patient welfare. As a countermeasure against this problem the Falsified Medicines Directive (FMD) 2011/62/EU should be operational until 2019 within all member states of the European Union. The main purpose is to protect patients by reducing the risk of counterfeits entering the supply chain. Therefore, an anti-counterfeiting system based on product serialization will be implemented. Each drug package will be assigned a unique identifier (2D barcode) and secured by a tamper-proof seal. This enables to track and verify each drug package along the supply chain from the manufacturer to the consumer. As a drawback, a central database managed by the European Medicines Verification Organisation (EMVO) is required. Manufacturers need to register new packages at the EMVO and pharmacies have to check-out each sold package. Actually, it is planned that additional costs are covered by the manufacturers but it is likely that those are passed to the consumers. Finally, a centralized system is exposed to getting compromised by forgers, e.g. by entering 2D barcodes from forged packages.

An alternative to serialization is packaging authentication based on classification which is inspired by physical object identification approaches relying on the concept of physically unclonable functions (PUFs). A PUF is a mapping between a challenge and response function which depends on the physical nature of a object. By definition a PUF is unique and cannot be reproduced. Related to packaging authentication various works dealed with Paper PUFs. Paper PUFs either rely on extrinsic or intrinsic PUFs, i.e. which are attached to the product or can be derived from a part of the product itself. However, PUFs are intended to identify an object. In case of classification-based authentication, it is assumed that the packaging of a product shows constant but discriminative intrinsic features. Instead of identifying each single package instance, it can be classified if the product is packaged with a specific packaging material or not. The focus in our research is on drug pills which are packaged in a blister and housed in a cardboard. Recently, in [3, 8] we investigated the basic feasibility of drug packaging authentication. In [3] we showed that cardboard textures of 9 different drugs from 3 manufacturers can be classified with 100% accuracy in a closed multi-class scenario. The utilized dataset was fairly small and packaging material authentication is in fact a simplistic binary classification problem, i.e. a single class has to be distinguished from all other classes. For the training stage only a limited subspace of known other classes is available which is referred to as open-set recognition. Thus, in [8] we focused on the open-set recognition problem and we investigated two basic pre-requirements for classification-based drug packaging authentication: positional invariance and instance generalisation of the packaging material texture. Based on a substantial database, with images of 45 different drugs from multiple instances (packages), both pre-requirements were proved successfully. However, all images were taken with a DSLR camera in an optimal setting and such imagery will not be available in case of a mobile device based authentication system. Thus, for this work in addition to a DSLR camera two smartphones were used to acquire a substantial dataset.

Based on this dataset in this work mobile-sensor as well as cross-sensor drug packaging authentication is investigated. Furthermore, in [8] only the particular classification accuracies for different parts of the packaging material were presented. For an authentication system it is assumed that the fusion of the particular classification

IH&MMSec '18, June 20–22, 2018, Innsbruck, Austria
© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery.
ACM ISBN 978-1-4503-5625-1/18/06…$15.00
https://doi.org/10.1145/3206004.3206016

results will increase the overall accuracy. Based on a simple majority voting approach, in this work the impact of fusion as well as feature selection will be elaborated. Finally, a closer look on possible authentication error sources will be presented. For example, it is assumed that parts of the packaging material from different drugs which are from the same manufacturer can be the same.

First, in Section 2 a possible scheme for a mobile device based drug packaging authentication system is introduced. Section 3 introduces the acquired database. The classification pipeline is outlined in Section 4. Experiments and results are presented in Section 5 and Section 6 concludes this paper.

2 MOBILE DEVICE DRUG PACKAGING AUTHENTICATION SYSTEM

A schematic illustration for a mobile-device based drug package authentication system is illustrated in Fig. 1. In order to proof the authenticity of a given drug the consumer will be guided by a mobile application. First, the user needs to disassemble the drug and to capture the textures of the cardboard (CB) and the blister top (BT) and blister bottom (BB) side. These three textures of the packaging material are denoted as modalities. The captured images are denoted as I_{CB}, I_{BT} and I_{BB}. Additionally, the user is advised to take a picture of the product code (I_{PC}), e.g. the European article number or the barcode printed on the cardboard. However, the product number can be entered manually or the respective drug can be selected from a list too. These four images compose the authentication vector $\hat{AV} = (I_{CB}, I_{BT}, I_{BB}, I_{PC})$ which is processed by the authentication system. First, the textural images I_{CB}, I_{BT} and I_{BB} are preprocessed. Preprocessing includes segmentation of the textural area and enhancement of the textural pattern. Subsequently, from each preprocessed image one patch is extracted for which a feature descriptor is computed. The product code image I_{PC} is used to determine the product code. Based on the product code, the system selects the corresponding precomputed classification models M_{CB}, M_{BT}, M_{BB} from a model repository. If the required models are not available on the mobile device they could be requested from a remote repository. Based on the corresponding models M_{CB}, M_{BT}, M_{BB} for each feature vector $FV_{CB}, FV_{BT}, FV_{BB}$ a probability score P_{CB}, P_{BT}, P_{BB} between $[0, 1]$ is computed. The closer to 1 the more likely the given feature vector is from a real sample, the closer to 0 the higher is the probability that the feature vector was computed from fake material. Finally, a decision function $f(P_{CB}, P_{BT}, P_{BB}) = (v, p)$ needs to be defined, where $v \in \{1, -1\}$ gives the final authenticity vote of the authentication system and $p \in [0, 1]$ specifies a probability score for the final vote which is then presented to the user.

3 DRUG PACKAGINGS TEXTURE DATABASE

For this study the same database as used in [8] and additional data captured with two different smartphones was utilized. Therefore, a Samsung S5 Mini & an IPhone 5 were utilized to capture images for a set of selected drugs. Therefore, mainly drugs with more than four instances from various manufacturers were selected. The acquisition setup is illustrated in Fig. 2f. Same as for the DSLR camera, the smartphones were mounted on a tripod and in addition a macro lens was utilized. For illumination a light source was placed laterally. An exemplary disassembled drug package is shown in Fig. 2a. The initial dataset consists of images from 45 drugs from 28 different manufacturers which were captured with a Canon 70D. For each drug between 1 and 15 package instances are available. The Canon 70D was mounted on a tripod and a 100mm lens and a flashlight were utilized (see Fig. 2e). From each drug instance images from the corresponding CB,BT&BB modalities were captured. For CB the inner side, showing the fibre structure was captured. For BT,BB the corresponding blister textures were captured. Thereby, it was ensured that the images were taken from different and non-overlapping regions. Examples depicting the variety of the different samples for each modality are shown in Fig. 2b-2d. All captured images were manually cropped ensuring that just texture remains. The images in the 1st row in Fig.3 illustrate exemplary images from each modality captured with the different sensors.

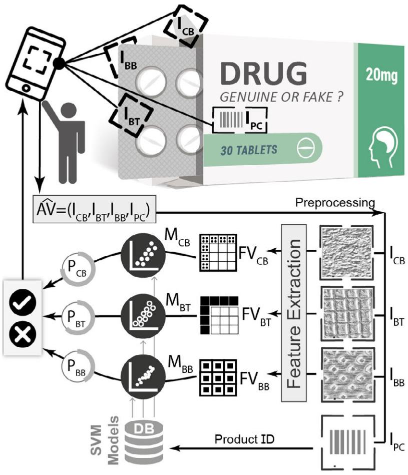
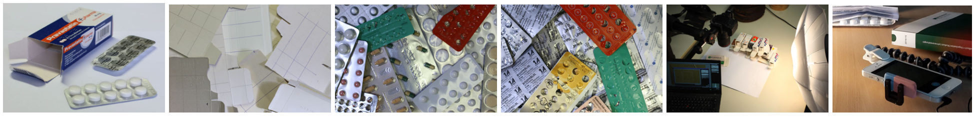


Figure 1: Mobile device drug packaging authentication

4 CLASSIFICATION PIPELINE

Data selection is essential for the subsequent cross-validation procedure. Due to the varying number of instances and the corresponding CB,BT&BB images per drug, a keypoint selection strategy has been employed. Therefore, a fixed number of data (k) to be sampled is predefined. Data relates to image texture patches of CB,BT&BB. For patch sampling, each CB,BT&BB image is subdivided into a grid which is specified by the size of the feature descriptor. According to the results presented in [8] 256×256 pixel patches are utilized. The 2nd row in Fig. 3 depicts sample images for CB,BT&BB for which the image patch grids are shown. Basically, k patches are selected from each instance of each drug and modality. However, k is only an upper bound of patches which are selected. For example, in this work k=1000 and especially for BT and BB there are drugs where less patches are available.

Image Enhancement. Prior to feature extraction the images are converted to grey-scale and Contrast Limited Adaptive Histogram Equalization (CLAHE) [10] is applied to each patch (parameters: block radius=50, bins=256, slope=40). Exemplary CLAHE enhanced



(a) Drug sample (b) Cardboards (CB) (c) Blister top (BT) (d) Blister bottom (BB) (e) Digital camera (f) Mobile camera

Figure 2: Image Acquisition Overview

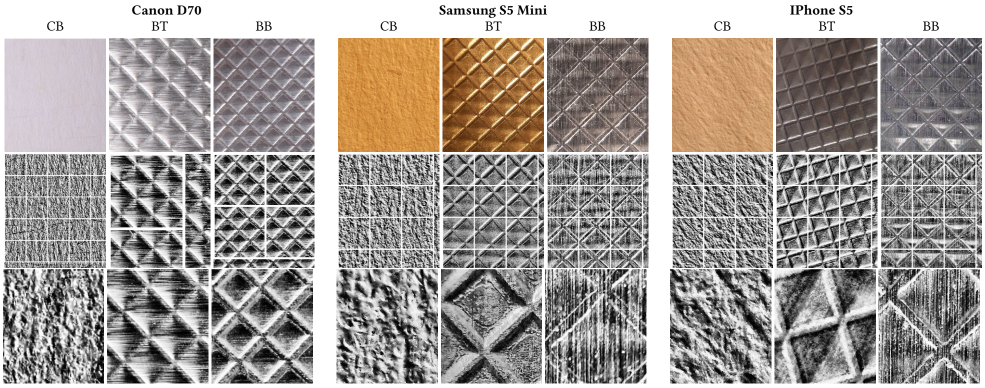


Figure 3: Preprocessing and data selection examples for Thrombo ASS produced by Lannacher Heilmittel (F1): 1st Row: Original images, 2nd Row: Preprocessed images showing the keypoint grid, 3rd Row: Exemplary 256×256 pixel patches from the top left keypoint in each image of the 2nd row.

images and selected patches for each modality and camera are shown in the 2nd and 3rd row of Fig. 3, respectively.

4.1 Feature Extraction and Feature Encoding

For each selected patch a feature vector using each of the following feature extraction approaches is computed: Local Binary Pattern (LBP) [5], Local Ternary Pattern (LTP) [9], Li Local Binary Pattern (LiLBP) [4], Speeded Up Robust Features (SURF) [1]. As noted in [8] IO and memory constraints are crucial when it comes to high dimensional features like SIFT and SURF. Furthermore, high dimensional feature vectors are computationally problematic in case of kernel-based SVM classifiers. As a first consequence the x,y step size for dense SURF method was increased to 16 pixel and we decided to compute both in a pyramid at three scales $(1, 2, 4)$. Consequently, for each patch #768 × SURF feature descriptors are computed. In case of SURF this results in a feature vector dimension of 98304. In preliminary tests it turned out that this feature vector size is suited for the classification experiments if a linear SVM classifier is utilized but not applicable in case of kernel SVMs.

Furthermore, image classification research showed that feature vector encoding schemes are beneficial for the classification accuracy. In case of SURF it was shown that the fisher vector (FV) encoding scheme [6] combined with linear classifiers improves the classification performance. The FV scheme encodes a set of vectors into a single vector which is composed by the first and second order residuals of the vectors from a Gaussian mixture model (GMM). Basically, the dimensionality of the fisher vector output is $2 \times K \times D$. K is the number of GMM components and D gives the feature vector dimensionality. Commonly, the FV encoding scheme is combined with a dimensionality reduction approach like Principial Component Analysis (PCA). Thereby, PCA is used to reduce the size of a feature vector to a predefined number of principal components. For this work, the input feature vector is reduced to 80 components. For a reduced input feature vector dimensionality of $D = 80$ and $K = 256$ Gaussian components a single FV with the size of $2 \times 80 \times 256 = 40960$ is produced. In case of SURF the FV encoding reduces the dimension of the SVM input vector by more than the half.

4.2 Data partitioning

In order to provide reliable results cross-validation (CV) based classification is performed. For each drug a number of instances (=packages) from each modality is available. Thus, a nested leave-one-package-out (LOPO) CV procedure is well suited to avoid overfitting and to force the computation of unbiased evaluation results.

The acquired database is composed by a set of drugs $D = \{d_1, ..., d_{45}\}$ produced by different $DM = \{dm_1, ..., dm_{28}\}$ drug manufacturers. $fdm(d_i) : D \rightarrow DM$ specifies the drug manufacturer for each drug. $M = \{CB, BT, BB\}$ specifies the packaging modalities. Furthermore the drugs and modalities were captured with different sensors $S = \{CANON = S1, SAMSUNG = S2, IPHONE = S3\}$ and different feature extraction methods $FE = \{fe_1, ..., fe_n\}$ are utilized in the experiments. The feature vector sets for a certain drug $d \in D$ and modality $m \in M$, for the k-patches from sensor $s \in S$ computed with feature extraction method $fe \in F$, are given by $FV_{(d,m,s,fe)} = \{fv_1, ..., fv_k\}$.

For binary classification it is required to specify a target class, i.e. the drug and the corresponding modality which we want to authenticate. In the scope of this work various classification configurations (CCs) are computed for each target drug d which are given by the following tuple: $CC = (d \in D, m \in M, s \in S, fe \in FE)$. The respective set of feature vector sets for a CC is given by $FV_{CC} = \{FV_{(d_1,m,s,fe)}, ..., FV_{(d_{45},m,s,fe)}\}$ which is composed by the CC specific feature vector sets from each drug. The positive training data $P_{CC} = FV_{(d,m,s,fe)}$ is specified by the target drug d in CC. The negative training data $N_{CC} = \{FV_{CC}\} \setminus \{FV_{(d,m,s,fe)}\}$ is composed by all feature vector sets of all other drugs. The positive and negative training data P_{CC}, N_{CC} are then used for nested cross-validation using a SVM classifier.

4.3 Cross-validation strategy

The overall goal of the CV strategy is to avoid two different types of over-fitting. The first ensures that no training data is used for evaluation as this leads to overestimation of the classification accuracy. CV excludes this type of over-fitting. The second type of over-fitting is crucial and concerns the training of the model. Thereby, hyperparameter selection plays a significant role in case of SVMs. The overall goal is to find parameters for a model which generalizes to the evaluation data, i.e. the ability of the model to classify unseen data. However, in binary open-set classification and especially in case of the considered drug authentication problem optimization is a trade-off between over- and under-fitting. Unseen data is composed by known data from the target drug and all other known drugs as well as a large set of data from unknown drugs. If the model is over-fitted to the training data it is likely that unseen evaluation data from other packages of the target drug are not recognized. On the other-hand under-fitting increases the risk that unseen as well as unknown packages from other drugs are misclassified as being the target drug.

Basically, for CV the positive and negative training data P_{CC} and N_{CC} for a certain CC are provided as input. For the LOPO CV strategy P_{CC} is split into n-folds $\{P_1, ..., P_n\}$ where each fold contains the feature vectors from a certain instance (=drug package sample). Thus, the number of folds n is given by the number of instances for the target drug d in CC which are available in the database. Same as in [8] the negative training N_{CC} data is split into known negatives KN_{CC} and unknown negatives $UN_{CC} = N_{CC}/KN_{CC}$. Therefore, for KN_{CC} the feature vector sets from a fixed number of drugs are selected, where the manufacturers are different to the target drug manufacturer of d in CC. The aim of this procedure is to simulate the real world, where only a limited set of other known drugs (faked and original ones) are available to train a classifier.

For the nested CV strategy in the outer loop we iterate over the n positive training folds. The current loop index is given by the variable i. In each iteration for KN_{CC} the features are split into two folds KN_1, KN_2 packagewise for each of the contained classes. Hence, half of the packages and the corresponding feature vectors of each class are contained in each fold. Subsequently, the ith positive and 2nd negative fold is selected for evaluation. The evaluation set is given by $E_{i,2} = P_i \cup KN_2 \cup UN_{CC}$. The unknown drugs UN_{CC} are only used for evaluation. The training set is composed by $T_{i,1} = \{P_1, ..., P_k\} \setminus \{P_i\} \cup \{KN_1\}$. Preliminary, $\{KN_1\}$ is reduced to a fixed number of feature vectors which are sampled equally distributed from all contained drug classes (=6) and the respective instances. In the inner CV loop for each $T_{i,1}$ the best hyperparameters are determined using a grid search approach. Same as in the outer loop, k-fold validation is performed repeatedly in order to test a set of SVM parameters. For this purpose, the known negative training data in $T_{i,1}$ is split classwise into two folds TKN_1 and TKN_2 (training known negatives). One fold simulates known negatives (=3 classes) and the other one unknown negatives (=3 classes) in the inner loop. While the known negatives are further used for training as well as for validation, the unknown negatives are just used for validation. It is assumed that this strategy is beneficial for the generalisation of the classifier. Hence, in the grid search procedure hyperparameters delivering a good classification accuracy in terms of the target class as well as known and unknwon classes accuracy are prioritized. As a measure for the performance the F-Measure is utilized which is well suited to balance between specialisation and generalisation in binary classification tasks. The utilized SVM classifiers assign each prediction a probability. In the inner loop, the probabilities are used to determine a threshold which maximizes the F-Measure. The SVM parameters and threshold delivering the highest F-Measure are selected for the outer loop. Those are then used to train and evaluate a classifier with the training and evaluation data from the outer loop, respectively.

5 EXPERIMENTS

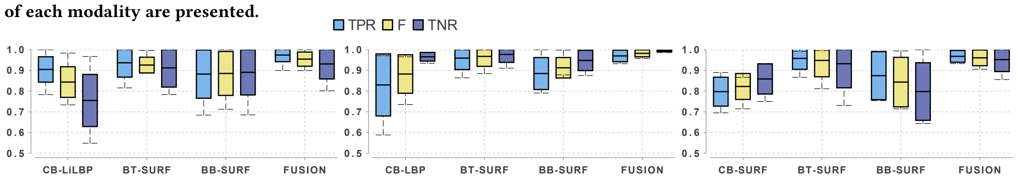
For data selection at maximum k=1000, 256× 256 pixel patches were selected from each modality and sensor. For each patch feature vectors are computed with all features listed in Section 4.1. In the experiments the LIBSVM linear SVM and kernel SVM with a radial basis function are utilized as classification approaches. Both are applied in combination with FISHER feature vector encoding (FVE=FISHER) and without (FVE=NULL) to cross-validate all CC combinations. Basically, the employed CV strategy requires that only drugs with at least 5 instances can be selected as target drugs, ie. the drug which should be authenticated by the classifier. An overview on suited drugs is presented in Table 2. The table shows that for each selected target drug various numbers of instances are available and each was captured with a set of sensors (S1,S2,S3). For each target drug and sensor all CCs are computed using the outlined LOPO CV strategy. For each LOPO CV the positive data is split into 2-folds, in the inner and outer CV loop. 6 drugs are selected for the known negative training data KN$_{CC}$. In order to assess the cross-sensor scenario, for evaluation in the outer CV loop data from all different sensors are utilized. For training data from only one sensor are utilized. For example, in case of Mexalen (A3) in the outer loop in each LOPO iteration the evaluation is performed with #1.75k-2k features of the target drug and >#100m features from all other drugs and cameras. For a fair evaluation of the different classification approaches and features the data splits are stored and reused.

5.1 Single-sensor evaluation

An overview on the particular results for the different sensors, all modalities and classification approaches is presented in Table 1. For each CC and modality the averaged results over all target drugs (Table 2) are shown. Considering the results for different CCs, it can be concluded that the F-Measure differences between the elaborated classifiers are not significant. For L-SVM and FISHER encoding it

CC		Canon - S1			Samsung - S2			IPhone - S3		
FVE	CA	CB	BT	BB	CB	BT	BB	CB	BT	BB
NULL	RBF-SVM	*LTP* 0.87 ±6.9	*LTP* 0.94 ±3.5	*LiLBP* 0.84 ±17.6	*LTP* 0.92 ±6.8	*LTP* 0.96 ±4.0	*LiLBP* 0.91 ±5.8	*LBP* 0.83 ±6.1	*LTP* 0.95 ±6.5	*LTP* 0.88 ±8.1
	L-SVM	*LTP* 0.87 ±7.4	*LBP* 0.92 ±4.7	*LiLBP* 0.83 ±13.5	*LTP* 0.92 ±6.3	*LTP* 0.94 ±4.1	*LiLBP* 0.9 ±5.6	*LBP* 0.83 ±6.9	*LTP* 0.95 ±6.2	*LTP* 0.8 ±12.6
FISHER	L-SVM	*LiLBP* 0.84 ±7.4	*SURF* 0.93 ±3.8	*SURF* 0.89 ±10.6	*LBP* 0.88 ±9.3	*SURF* 0.97 ±4.8	*SURF* 0.91 ±4.9	*SURF* 0.82 ±6.3	*SURF* 0.95 ±7.9	*SURF* 0.84 ±12.0

Table 1: Single-sensor performances: For each sensor and all CCs the mean F-Measure and the StDev[%] for the best features of each modality are presented.



(a) S1 - Canon

(b) S2 - Samsung

(c) S3 - IPhone

Figure 4: Single-sensor results for FISHER L-SVM: For each sensor and modality the performances for the best features as well as for modality fusion are depicted. TPR = $\frac{TP}{TP+FN}$, TNR = $\frac{TN}{TN+FP}$ [Y-Axis: Mean, min, max, standard deviation].

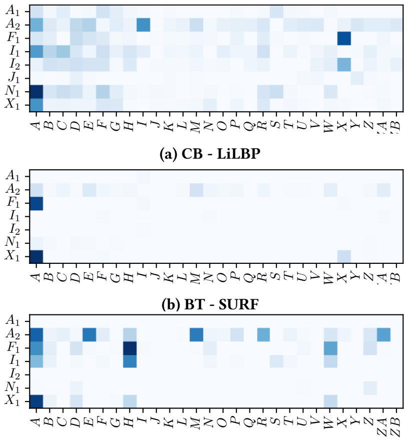
seems that SURF as high level feature does not improve the performance as expected. Furthermore, the F-Measures are comparable to the results presented in [8]. However, in [8] less data was selected for training which shows that doubling the parameter k to 1000 does not improve the classification performance.

When comparing the F-Measures between the different sensors the values are in the same range, surprisingly. Basically, for the mobile sensors fewer drugs were available for evaluation, i.e. no unknown drugs remain for evaluation. Thus, it would be assumed that less variety (=closed-set) in the evaluation data improves the classification performance. This new finding is interesting because this increases the chance that the classification performances are robust in a real world application.

Modality fusion. In the experiments in [8] only the modality performances were considered. As shown in the exemplary drug packaging authentication scheme in Fig. 1 the three probability scores from each modality (P_{CB}, P_{BT}, P_{BB}) should be combined to a final decision. For this purpose, a simple majority voting approach is applied which still offers possibilities for optimization. Initially, the modality specific classifier thresholds are used to determine a decision vector $\hat{D} = (D_{CB}, D_{BT}, D_{BB})$ from the probability scores. The decision values are either 1 or -1. In case that at least two decision values are 1 the final decision is that the package material is from a real package, i.e. it is not a fake sample. For the selection of the features which achieve the highest F-Measure SFFS (Sequential Floating Forward Selection) [7] is applied. For this purpose, the particular modality decisions are randomly shuffled to to get a set of decision vectors. The shuffling is repeated several times in order to compute the averaged classification performances of the modality fusion. For each sensor the particular modality performances as well

Manufacturer/Drug	#Samples		Camera		
	CB	BT&BB	Canon (S1)	IPhone (S2)	Samsung (S3)
(A) ratiopharm					
(A1) Danselle	10	10	✓	-	-
(A2) Danseo	9	9	✓	-	-
(A3) Mexalen	8	8	-	✓	✓
(F) Lannacher					
(F1) Thrombo ASS	5	5	✓	✓	✓
(I) Kwizda Pharma					
(I1) Liberel mite	15	15	✓	-	-
(I2) Delia	11	11	✓	✓	✓
(J) Rotexmedia					
(J1) Dexamethason	5	0	✓	-	-
(N) Gynial					
(N1) Bilinda	6	6	✓	✓	✓
(X) Pelpharma					
(X1) Peliette	17	17	✓	✓	✓

Table 2: List of drugs with at least 5 instances which were selected as target drugs. Only drugs which were captured with the corresponding sensors show a check-mark.

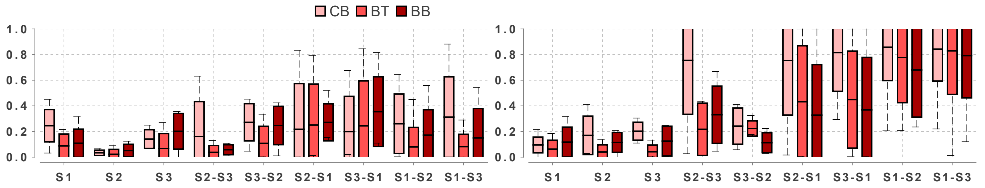


(a) CB - LiLBP

(b) BT - SURF

(c) BB - SURF

Figure 5: Single-sensor results for Canon (S1) FISHER L-SVM: (FN+FP) Error matrix for each modality. [X-Axis: Producers from the evaluation data, Y-Axis: Target Drugs]. The darker the cell, the higher is the classification error.



(a) False positve rates: FPR=$\frac{FP}{TP+FP}$

(b) False negative rates: FNR=$\frac{FN}{TN+FN}$

Figure 6: Cross-sensor performances for FISHER L-SVM: For all training and evaluation sensor combinations the FPR and FNR for each modality are shown. For each combination and modality the results for the best feature were selected. [Y-Axis: FPR/FNR mean, min, max and standard deviation].

as the fusion performance is illustrated in Fig. 4. It can be concluded that modality fusion significantly improves the classification and authentication accuracy.

Error sources. Basically, it is assumed that other drugs from the same or different manufacturer might have the same packaging material, e.g. if two different manufacturers have the same cardboard or blister supplier. The error matrix plots in Fig. 5 visualize the number of false positive (FP) + false negative (FN) votes for each target drug and the evaluated drugs which are grouped into manufactures. The darker the higher the amount of misclassification's. FP votes are from samples which are incorrectly authenticated and FN votes are from samples which were incorrectly not authenticated. When considering the columns it can be observed how likely the drugs of a certain manufacturer cause FP or FN votes. FN votes are only possible when the target drug (e.g. A1) and the manufacturer (A) in the columns are the same. For example, for all three modalities the drugs of ratiopharm (A) cause FP votes for drugs from other manufacturers as well as FN votes for A1 and A2. Furthermore, each target drug and the corresponding row can be considered. The darker the more FP and FN votes were observed in the CV strategy. In case of CB, the drug A2 shows a high amount of errors. Furthermore, in each error matrix there are some dark spots which show up high error rates. For example, for BB a high amount of samples from manufacturer H are incorrectly classified as drug F1 = FP votes. Comparing the error matrices for all three modalities it is obvious that the most errors are visible in case of CB and BB and there are less errors for the BT textures.

5.2 Cross-sensor evaluation

In order to assess the cross-sensor performances, all CCs were evaluated with data from other sensors. Thereby, the classifier was always trained with data from only one sensor. The two charts in Fig.6 show the FP and FN rates which were achieved for different training and evaluation sensor combinations. Actually, S1,S2&S3 show the single sensor FPR and FNR for each modality. All other combinations show results where the classifier has been trained with data from one sensor and has been evaluated with data from another sensor, i.e. cross-sensor results. The single-sensor error rates are in general lower than the cross-sensor results for almost all modalities. Especially, the cross-sensor combinations where either the DSLR or a mobile camera are used for training and the other camera type is used for evaluation show inferior FNR values and also worse FPR values. This could be attributed to the different texture scales in case of images acquired with the DSLR camera and images acquired with the mobile devices (see Fig. 3). Backing for this argument is that the error rates for the mobile-device cross-sensor combinations are better. Furthermore, the cross-sensor FNR values are inferior to the FPR values compared to the single sensor results. Thus, in the considered cross-sensor scenario it is easier for the classifier to reject samples from other drugs than to detect samples from the same drug captured with a different sensor.

6 CONCLUSION

In this work different aspects for a mobile device based drug packaging authentication system were considered. Results showed that data captured with mobile devices and low level features are principally suited for drug packaging authentication. Furthermore, modality fusion improves the performance significantly. However, if different sensors are used and the imaging conditions get more realistic the authentication performance degrades significantly.

Future work on a mobile device based application needs to deal with all issues caused by unconstrained imaging conditions (scale, rotation, tilt & illumination variations). Furthermore, more sophisticated approaches for modality fusion, state-of-the art features and a CNN-based solution should be employed.

REFERENCES

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. 2008. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* 110 (June 2008), 346–359. Issue 3. https://doi.org/10.1016/j.cviu.2007.09.014

[2] EUIPO. 2016. The economic cost of IPR infringement in the pharmaceutical industry. http://authenti-city.eu/wp-content/uploads/2016/10/The-Economic-Cost-of-IPR-Infringement-in-the-Pharmaceutical-Industry-EN.pdf. (2016).

[3] Christof Kauba, Luca Debiasi, Rudolf Schraml, and Andreas Uhl. 2016. Towards Drug Counterfeit Detection Using Package Paperboard Classification. In *Advances in Multimedia Information Processing – Proceedings of the 17th Pacific-Rim Conference on Multimedia (PCM'16)* (September 15 - September 16) *(Springer LNCS)*, Vol. 9917. Xi'an, CHINA, 136–146. https://doi.org/10.1007/978-3-319-48896-7_14

[4] Z. Li, G. Liu, Y. Yang, and J. You. 2012. Scale- and Rotation-Invariant Local Binary Pattern Using Scale-Adaptive Texton and Subuniform-Based Circular Shift. *IEEE Transactions on Image Processing* 21, 4 (April 2012), 2130–2140.

[5] T. Ojala, M. Pietikäinen, and T. Mäenpää. 2002. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 7 (July 2002), 971–987.

[6] F. Perronnin and C. Dance. 2007. Fisher Kernels on Visual Vocabularies for Image Categorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07).* 1–8.

[7] P. Pudil, J. Novovicova, and J. Kittler. 1994. Floating Search Methods In Feature-Selection. *Pattern Recognition Letters* 15, 11 (November 1994), 1119–1125.

[8] Rudolf Schraml, Luca Debiasi, Christof Kauba, and Andreas Uhl. 2017. On the feasibility of classification-based product package authentication. In *IEEE Workshop on Information Forensics and Security (WIFS'17).* Rennes, FR.

[9] Xiaoyang Tan and Bill Triggs. 2007. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. In *Analysis and Modelling of Faces and Gestures (LNCS)*, Vol. 4778. 168–182.

[10] K. Zuiderveld. 1994. Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems IV*, Paul S. Heckbert (Ed.). Morgan Kaufmann, 474–485.

Near-Infrared Illumination Add-On for Mobile Hand-Vein Acquisition

Luca Debiasi, Christof Kauba, Bernhard Prommegger and Andreas Uhl
University of Salzburg
Jakob-Haringer-Str. 2, 5020 Salzburg, AUSTRIA
{ldebiasi, ckauba, bprommeg, uhl}@cs.sbg.ac.at

Abstract

There is a growing need for mobile authentication solutions. Biometric recognition systems provide several advantages over conventional knowledge and token based solutions. Especially the use of vascular patterns as a biometric trait gains more and more attention. We present a near-infrared illumination add-on for smartphone devices which allows to capture the vascular pattern of the hands (hand-veins). This device is connected and controlled via Bluetooth and customised for the Nexus 5 smartphone but can be easily adopted to fit other models too. Due to the inherent risk of fraudulent authentication attempts on a non-trusted platform like a smartphone, we propose a challenge-response approach to ensure the authenticity of the captured hand-vein images. A hand-vein data set comprising of 31 subjects and 920 images in total is acquired with the presented device. A performance evaluation utilising different hand-vein recognition schemes is conducted to show the applicability of our device and the proposed challenge-response approach.

1. Introduction

Mobile authentication solutions enjoy a wide-spread use nowadays. No matter if for payment transactions, unlocking a mobile device or identity verification at border control, there is a growing need for mobile authentication solutions. Especially the application of biometric recognition technologies in the scope of mobile authentication is gaining more and more attention. Biometrics provide several advantages over traditional means of authentication in terms of resistance against forgery and user's convenience. Fingerprint recognition systems have been integrated into higher class smartphones (e.g. the Samsung Galaxy S6 and onwards, the Apple iPhone and several more) for several years now and also face as well as iris recognition systems find their way to the newest generation of smartphones (e.g. in the Samsung Galaxy S8/S8+). Beside these traditional biometric traits, vascular pattern based ones have become an emerging biometric trait during the last years. Vascular pattern based recognition (commonly denoted as vein recognition) can help to overcome some of the problems existing biometric recognition systems have. Vein based systems rely on the structure of the vascular pattern formed by the blood vessels inside the human body tissue. This pattern only becomes visible in near-infrared (NIR) light. Thus, vein based biometrics provide a good resistance to spoofing and are insensitive to abrasion and skin surface conditions. They achieve a competitive recognition performance while the user http://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2010.0191convenience is at the same level as for fingerprint systems as long as the scanner is designed in an open manner. Moreover, a liveness detection can be performed easily [11] and a contactless operation is possible, which is especially important for mobile authentication solutions. This makes vein pattern based systems a valuable choice in the scope of mobile authentication.

The application of biometric recognition systems in mobile scenarios rises some problems compared to the stationary use of these systems. First of all, the acquisition process is more unconstrained (more degrees of freedom for the placement of the biometric and varying environmental conditions) compared to the stationary case, causing several recognition performance issues [4, 5, 17]. Second, the authentication process is unsupervised, enabling presentation attacks [1, 10]. Furthermore, the mobile system might not be a trusted platform, especially if the authentication is performed on the user's smartphone. This opens the door for all kinds of insertion and replay attacks to the biometric system. Hence, there is the need for presentation attack detection systems as well as methods to prove the authenticity and integrity of the biometric sample that has been captured.

In this work we present a smartphone add-on to acquire hand-vein images. In contrast to other mobile vein scanner solutions in the literature, our add-on module is basically an illumination module only, lowering its production costs compared to full-fledged scanner devices. It utilises the phones integrated camera to capture the vein images and is controlled wirelessly via Bluetooth by our custom designed capturing Android app. Thus, together with a suitable smartphone, this add-on resembles a full mobile hand-vein scanner. Unlike most previously proposed mobile scanners, our device operates fully contactless without a specifically designed device to keep the hand in a predefined position.

To cope with the inherent risks of insertion attacks, the capturing process features a challenge-response protocol based on varying illumination intensities. In this way the app is able to prove that an actual image of the vein patterns has been captured and no previously captured sequence has been inserted instead. We established a mobile hand-vein data set captured with our mobile hand-vein add-on in combination with a modified Nexus 5 smartphone that will be made publicly available. It comprises 31 subjects and 920 images in total. Based on this data set a performance evaluation using several well-established hand-vein recognition schemes is conducted in order to show the decent recognition performance that can be achieved using our mobile hand-vein scanner and to prove the effectiveness of the challenge-response approach.

The rest of this work is organised as follows: Section 2 gives an overview of related work on mobile finger- and hand-vein scanners. The details of our proposed mobile hand-vein scanner add-on, including the challenge-response protocol, and the differences to previous mobile vein scanners are described in Section 3. Section 4 presents the publicly available data set that has been captured with our hand-vein scanner add-on. Section 5 deals with the performance evaluation. At first the details of the employed hand-vein recognition tool-chain are described, followed by the evaluation results and a results discussion. Section 6 concludes this paper and gives an outlook on future work.

2. Related Work

This section gives an overview on related work in mobile and embedded finger- as well as hand-vein scanner devices. Liu et al. [13] proposed a “a real-time embedded finger-vein recognition system for authentication on mobile devices”. Their scanner consists of an NIR sensitive monochrome camera with an additional NIR pass-through filter, a white acrylic plate where the finger is placed onto and a NIR laser based illuminatior below this plate (light transmission principle). They equipped the NIR lasers to cope with problems due to shadows caused by the LED light source within their scanner design. Their full-fledged recognition system is implemented on a DSP (digital signal processor) and features image acquisition, ROI (region of interest) extraction, feature extraction and comparison. The DSP integration enables a mobile application. Sierro et al. presented three prototype touch-less vein scanners, a finger- and two palm-vein ones in [19]. The touch-less nature makes this system more convenient for the user and less susceptible to spoofing. All their proposed scanners are based on the reflected light principle (illumination source and camera on the same side of the hand/finger). Their first palm-vein prototype contains a Sony ICX618 659x494 CCD camera together with a 920 nm long-pass filter, 20 940 nm NIR LEDs and an ultrasonic sensor to detect the distance between the scanner and the hand. Their second palm-vein prototype features multi-spectral acquisition to increase its robustness against simple types of spoofing attacks by equipping blue and far-red LEDs in addition to the NIR ones. The layout of the LED positioning was changed too, but all other components are the same as within their first prototype. The finger-vein prototype consists of a OV7670 Color 640x480 pixel CMOS sensor in combination with a wide angle 2.1 mm lens and an infrared long-pass filter with a low cut-off wavelength of 740 nm. The 12 NIR LEDs are arranged in three groups of 4 LEDs each to enable an optimal illumination of the finger-veins. All three proposed scanners are small in size. The finger-vein one is USB-host powered and can be controlled by an Android app which facilitates its use as mobile finger-vein scanner. The palm-vein ones can be modified to be USB-host powered and work in combination with a smartphone too.

Eng and Khalil-Hani proposed several versions of a FPGA-based vein biometric authentication system. In [3] they introduced an embedded hand vein scanner implemented on an Altera Nios II prototyping system running on Nios2-Linux as real time operating system. Their sensor consists of an reflected light source (NIR LED array), a modified thermal webcam with a resolution of 320x240 pixels and an attached IR filter. They captured images from the dorsal (back) side of the hand and utilised minutiae based features extracted from the vein pattern for recognition. In [8] and [9] they proposed two versions of a finger vein recognition system. Again, the system was implemented on an Altera prototyping system running Nios2-Linux using a modified webcam. Contrary to the hand vein scanner, they used the light transmission method to acquire images from the palmar side of the finger. For recognition they use minutiae based methods again.

Lee et al. presented a mobile multimodal biometric capture device utilising finger-veins and fingerprints in [12]. Their scanner consists of two QuickCam USB cameras, a visible light source for fingerprints and four 880 nm NIR LEDs for finger-veins using the light transmission principle. The captured images have a resolution of 640x480 pixels. Their embedded system unit is a ultra mobile processing computer manufactured by SONY Corp (VGN-UX17LP). They used a minutiae based recognition method for both, finger-veins and fingerprints.

Fletcher et al. proposed two mobile hand-vein scanners in [4]. The first one uses an unmodified Sony Experia

Mini Pro Android smartphone as camera. The light source consists of 16 NIR LEDs with an operation frequency of 850 nm. They used a Kodak Wratten filter (#87) with a pass-through frequency range of 740-795 nm as optical filter. The second one uses a Gearhead Nightvision webcam (WC1100BLU) which already contains six IR LEDs as light source. For a better illumination they replaced the internal LEDs with 940 nm NIR LEDs. Again, they used a Kodak Wratten filter (#87c) with a pass-through frequency range of 790-855 nm as optical filter. The webcam is attached to a Nexus 7 Android tablet. Both scanners used especially designed apparatuses to place the hand into a well defined position. They acquired the vein structure from the palmar side and used minutiae based features for biometric recognition.

In contrast to existing mobile hand- and finger-vein scanners, our proposed mobile hand-vein scanner is basically an illumination add-on module for smartphones. Thus, it exhibits lower production costs compared to a full featured scanner device.

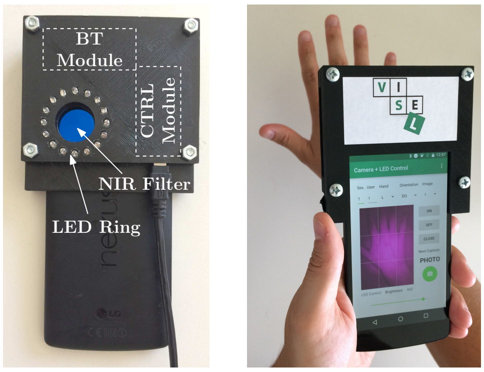


Figure 1: Left: mobile hand-vein scanner add-on, right: typical acquisition set-up

3. Mobile Hand-Vein Scanner Add-On

As mentioned in Section 2, the basic components of a hand-vein scanner are an NIR light source and an NIR-sensitive image sensor (camera). Every common smartphone nowadays has a built-in camera, which is sufficient to capture high-resolution hand-vein images. However, these cameras are usually equipped with NIR blocking filters in order to avoid unwanted colour effects in the captured images. Thus, it is necessary to either remove this NIR blocking filter, like it has been done for the modified Nexus 5 smartphone by EigenImaging (`https://www.eigenimaging.com/`) we utilised, or a separate NIR-sensitive camera has to be equipped in the smartphone. The latter is done by some manufacturers already (e.g. Samsung for iris recognition) and it is likely that others will follow this trend. If the smartphone already contains an NIR-sensitive camera, it can be utilised for hand-vein recognition and the only additional component needed is the NIR light source.

Our mobile hand-vein scanner add-on is essentially such a light source in the form of an add-on module for smartphone devices, depicted in Fig. 1. It is the first mobile hand-vein scanner device of its kind, exhibiting lower production costs compared to previous mobile finger- and hand-vein scanners as it does neither contain a separate image sensor nor a complex control board. The whole device was designed to be used in combination with our modified Nexus 5 smartphone utilising its integrated camera as image sensor and has been constructed by ourselves. The housing part, where the smartphone is slid in, consists of several, 3D printed components. Hence, it can be easily modified for other smartphone models. This control board is based on an Arduino Nano Board (`https://store.arduino.cc/arduino-nano`), a Bluetooth module and a 16 channel LED driver IC. The control board design was adopted from our previous finger-vein scanner [6] and modified for the mobile application. It does not require a physical cable connection to the smartphone, the data transfer and control is achieved via Bluetooth communication. The USB cable on the prototype is only needed for power supply, but the final version will have a built-in rechargeable battery. It has 16 NIR LEDs with a peak wavelength of 850 nm that are arranged in a circle around the smartphone's camera. Each LED can be brightness controlled individually. This enables a uniform and sufficient illumination for the hand-vein images on the one hand and provides the ability to use complex illumination patterns for encoding information on the other hand. Moreover, our add-on is equipped with an NIR pass-through filter having a cut-off frequency of 780 nm to filter out the ambient light and to improve the image contrast.

The LED brightness is automatically controlled by a capturing app running on the smartphone, prior to the acquisition of a single image. Currently, the app only supports the capturing of single images as well as video sequences of the hand-veins, no feature extraction and comparison is done yet. This is why it is currently an add-on for the acquisition of hand-vein images and not for performing a full authentication.

3.1. Challenge-Response Protocol

In order to prevent presentation and replay attacks, the developed mobile hand-vein scanner add-on is capable of performing challenge response (CR) authentication due to its 16 fully controllable NIR LEDs.

Presentation attacks for finger veins have already been successfully conducted in [21] by Tome et al., where a spoofing false accept rate of 85% has been achieved. Their experiments have been conducted using an extensible framework for spoofing finger veins, which might also be successfully applied to hand veins.

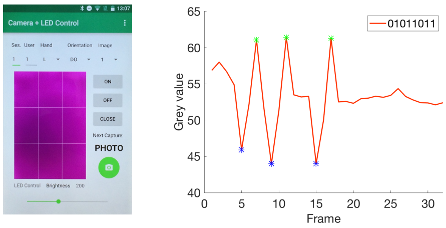
CR authentication follows the simple principle that one party presents a question, i.e. the *challenge*, to which another party has to provide a valid answer, i.e. the *response*, in order to pass the authentication. In [20] Stein *et al.* proposed a video-based fingerprint recognition and anti-spoofing solution for smartphones. They developed a CR protocol, where the finger needs to be moved towards the camera and the reflectance of the finger surface is measured.

In the scenario presented here, the first party (user) tries to authenticate itself in a biometric system using his/her smartphone. More precisely, the smartphone is used as a mobile sensor to acquire the user's biometric trait, i.e. the vascular pattern of the hand, and the data is then submitted to the biometric system wirelessly for the identification. Since the smartphone cannot be trusted, as mentioned in Section 1, the biometric system has to ensure the authenticity and up-to-dateness of the acquired biometric data to prevent a malicious insertion or spoofing of the submitted data, i.e. presentation and replay attacks. Therefore, a video of the hand is acquired for authentication, which contains a blinking sequence generated based on a specific challenge using the 16 fully controllable NIR LEDs. This blinking sequence is an inherent part of the video, which is sent to the biometric system after acquisition. This ensures that the response is interwoven with the biometric data.

The proposed challenge response protocol consists of the following steps:

1. The smartphone sends an authentication request to the biometric system.
2. The biometric system generates a random number which defines a fixed blinking sequence and sends it to the smartphone (challenge).
3. The smartphones generates a blinking sequence based on the random number and controls the 16 LEDs accordingly. In parallel, a video of the hand (dorsal or palmar) is recorded.
4. The video containing the biometric data and blinking sequence (response) is sent wirelessly to the biometric system.
5. The biometric system detects the blinking sequence and compares it to the previously generated random number.
6. If the response matches the challenge, the hand vein recognition is performed and the user is authenticated. Otherwise, the whole process is repeated.

We implemented this challenge response protocol in form of an Android application, which runs on the user's smartphone. The app consists of two major parts: The video recording and the LED control. The video recording part has been realised using *CameraView* (`https://github.com/natario1/CameraView`), a high-level library providing access to the smartphone's camera in order to capture photos and videos. The LED control is performed via Bluetooth. The application is able to capture both photos and videos and contains different settings to configure the acquisition parameters for testing and development. The GUI of the developed application is depicted in Figure 2a.



Ses. 1 User 1 Hand L Orientation DO Image 1

LED Control Brightness 200

(a) GUI

(b) Grey values

Figure 2: Graphical user interface (a) of the developed Android application for LED control and capturing of photos and videos. Sequence of mean grey values (b) for an exemplary video with detected 0s (marked as blue stars around a gray value of 45) and 1s (marked as green stars around a gray value of 60).

In our first proof of concept, we generate a random number between 1 and 255, which is logged to a file for later evaluation. This random number, which by concept would be sent by the biometric system, is then transformed into an 8-bit binary sequence: For 0 the brightness of the LEDs is reduced to half of its intensity, while for 1 the intensity is kept at a predefined level. All LEDs are controlled equally in the current version, but multiple illumination zones can be realised in the future to enable more complex blinking sequences. For this work, we acquired videos with a duration of 3 seconds containing the 8-bit sequence, leading to a blinking interval duration of 375 ms. For synchronisation purposes, we added a padding with a duration of 1.5 seconds before and after the blinking sequence of the video.

The biometric system, which receives the video, has only been simulated so far. Therefore, the blinking sequence is first extracted from the recorded video and compared to the previously logged random number for the specific video. For the detection of the sequence, single frames are extracted from the video at a frame-rate of 5.33 fps using FFmpeg (`https://www.ffmpeg.org`). The frame-rate has been selected in correspondence to the blinking interval of 375 ms, leading to two frames being extracted for each blinking interval and a total of 32 frames for the whole video: 8 images of padding before and after the blinking sequence and 16 images for the sequence itself. Thereafter, the mean grey value of each extracted image is determined. Since the used camera library does not allow a manual exposure, it is automatically regulated after each blinking to obtain a certain mean grey value. Hence, we are only able to detect changes in illumination. Afterwards, the local minima and maxima, i.e. 0s and 1s, of the blinking sequence are determined from the mean grey value curve, as shown in Figure 2b. The intervals between these extrema, i.e. where no illumination change has happened, are interpolated by the preceding value. With this procedure, we obtain an 8-bit binary sequence again which is matched against the 8-bit binary sequence of the random number.

4. Mobile Hand-Vein Data Set

The mobile hand-vein data set was acquired using our mobile hand-vein scanner add-on in combination with the modified Nexus 5 smartphone. It includes dorsal as well as palmar hand-vein images of 31 individual subjects. No supporting apparatuses to place the hand in a predefined positional were used. As a result, the captured images resemble a realistic real-live scenario with all possible types of distortions like rotation, tilting in all possible directions and scaling (different distances of the hand and the smartphone). The data acquisition was split into two separated sessions, the first one outdoor inside a car and the second one indoors. Throughout the first session 28 subjects have been acquired, during the indoor session 18. 15 subjects participated in both sessions. Five images per hand and per view have been acquired, summing up to a total of 920 images. The acquisition outside was done to simulate a realistic application scenario of our mobile hand-vein add-on in a border control environment, the inside session was conducted to have reference images in a more controlled environment. The acquired colour JPEG images have a resolution of 2448x3264 pixels. We extracted square ROI patches of the hand-vein images manually, which have a resolution of 512x512 pixels. Figure 3 shows some example images. This data set will be publicly available as part of the PROTECT Multimodal DB Dataset [22] database and can be downloaded at `http://projectprotect.eu/`.

5. Experimental Evaluation

In the following the finger-vein processing tool-chain and the evaluation protocol are described. Then the experimental results are given and discussed.

5.1. Processing Tool-Chain

The finger-vein processing tool-chain consists of ROI extraction, preprocessing, feature extraction and comparison. We opted for simple binarisation type feature extraction methods as well as two key-point based methods (one SIFT based and an adopted version of an algorithm proposed by Matsuda et al. in [15]) to have a complimentary feature type too.

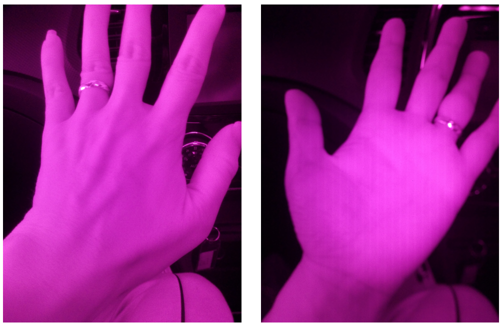


Figure 3: Example images of the mobile hand-vein data set, left: dorsal, right: palmar

ROI Extraction The ROI extraction is done manually by fitting a rectangular ROI is fit inside the hand area. The ROI images have a size of 512×512 pixels.

Preprocessing To improve the image contrast and the visibility of the vein pattern **CLAHE** [25], which is the most prevalent and simple technique, in combination with **High Frequency Emphasis Filtering (HFE)** [24] and filtering with a **Circular Gabor Filter** (**CGF**) as proposed by Zhang and Yang [23] are applied. Furthermore, the images are resized to half of its original size, which not only speeds up the comparison process but further improves the results due to intrinsic denoising. For more details on the preprocessing methods the interested reader is referred to the authors' original publications.

Feature Extraction and Comparison The first three of the following techniques aim to extract the vein pattern from the background resulting in a binary template image followed by a comparison of these binary templates using a correlation measure.

Maximum Curvature (MC [16]) aims to emphasise only the centre lines of the veins, making it insensitive to varying vein widths. The first step is the extraction of the centre positions of the veins. Afterwards a score according to the width and curvature of the vein region is assigned to each centre position and recorded in a matrix called locus space. Due to noise or other distortions some pixels may not have been classified correctly at the first step, thus the centre positions of the veins are connected using a filtering operation. Finally binarisation is done by thresholding using the median of the locus space.

Principal Curvature (PC [2]): At first the gradient field of the image is calculated. Hard thresholding is done to filter out small noise components and then the gradient at each pixel is normalised to 1 to get a normalised gradient field. This is smoothed by applying a Gaussian filter. The next step is the actual principal curvature calculation, obtained from the Eigenvalues of the Hessian matrix at each pixel. Only the bigger Eigenvalue, corresponding to the maximum curvature, is used. The last step is a binarisation of the principal curvature values to get the binary vein output image.

Gabor Filter (GF [11]): The image is filtered using a filter bank consisting of several 2D even symmetric Gabor filters with different orientations, resulting in several feature images. The final vein feature image is obtained by fusing all these single images, which is then post-processed using morphological operations to remove noise.

For comparing the binary feature images we adopted the approach of Miura et al. [16]. As the input images are neither registered to each other nor aligned vertically, the correlation between the input image and x- and y-direction shifted versions of the reference image is calculated. The maximum of these correlation values is normalised and then used as final comparison score.

In addition to the techniques described above, the fourth technique is a key-point based one. Key-point based techniques try to use information from the most discriminative points as well as considering the neighbourhood and context information of these points by extracting key-points and assigning a descriptor to each key-point. We used a **SIFT** [14] based technique with additional key-point filtering along the finger boundaries as proposed by Kauba et al. [7] and a modified version of **Deformation-Tolerant Feature-Point Matching** (DTFPM) proposed by Matsuda et al. [15]. DTFPM was designed for finger-vein recognition. Its feature extraction assumes a circular shape of the finger. This does not apply for hand-vein recognition, thus we modified the feature extraction step.

5.2. Evaluation Protocol

The experiments are split into two main parts: in the first part we analyse the recognition performance of the database. For evaluation purposes, dorsal and palmar images are regarded as two independent data sets. In addition to the analysis of the two acquired sessions, we performed a comparison of session 1 against session 2 as well. To quantify the performance, the EER as well as the FMR100 (the lowest $FNMR$ for $FMR <= 1\%$), the FMR1000 (the lowest $FNMR$ for $FMR <= 0.1\%$) and the ZeroFMR (the lowest $FNMR$ for $FMR = 0\%$) are used. We applied the following test protocol: For calculating the genuine scores, all possible genuine comparisons are performed. For calculating the impostor scores, only the first image of a finger is compared against the first image of all other fingers. Table 1 states the number of comparisons for each evaluation. As our recognition scheme does not require a training step, no separate training and test set is needed. All result values are given in percentage terms, e.g. 1.43 means 1.43%. In the second part of our experiments, we evaluated the captured videos with respect to the challenge-response protocol described in Section 3.1. A public implementation of the complete processing tool-chain as well as the scores and detailed results are available at: `http://www.wavelab.at/sources/Debiasi18b`.

	Session 1	**Session 2**	**Session 1 vs 2**
Genuine	560	360	750
Impostor	1540	630	855
Total	2100	990	1650

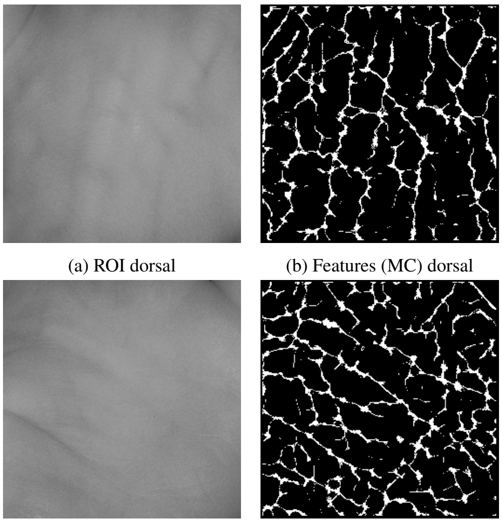
Table 1: Number of matches per data set/session evaluation

5.3. Recognition Performance Results

The performance evaluation has been conducted for both, the palmar and dorsal sub-set. Figure 4 shows some sample images including te extracted MC features. In the dorsal ROI image, the vein structure is visible. In the images acquired from the palmar side (bottom row), the vein structure is not visible as prominently. The ROI image on the left side is dominated by the texture of the palm. This fact is also reflected in the extracted MC features (right side): most of the extracted lines do not result from the vein structure but from the creases and wrinkles of the skin.

Table 2 lists the results for the dorsal subset. For session 1 (outdoor) MC achieves the best result with an EER of 4.13% followed by DTFPM (7.33%), SIFT (10.63%) and PC (10.71%). With an EER of 28.08%, GF perform significantly worse than all other feature types. For session 2 (indoor, controlled ambient light) all feature types except PC perform worse. MC still shows the best performance with and EER of 5.69%. PC (8.97%) now achieves a better result than DTFPM (12.00%) and SIFT (14.17%). Again, the recognition performance of GF (36.37%) is not competitive to the other methods. For the inter-session comparison, the performance drops dramatically. MC achieves an EER of only 24.30% which is six times worse than the result for session 1. PC and DTFPM exhibit EERs around 30%, SIFT and GF of greater than 40%. The DET plots for session 1 and 2 are depicted in Figure 5 left and right, respectively.

Table 3 states the results for the palmar sub-set. The results follow the same trend as for the dorsal sub-set: MC performs best for all 3 experiments followed by DTFPM, SIFT and PC. The outdoor session exhibits a better performance than the indoor session and the inter-session comparison performs significantly worse than the single sessions.



(a) ROI dorsal

(b) Features (MC) dorsal

(c) ROI palmar

(d) Features (MC) palmar

Figure 4: Sample images of both data sets: dorsal images on the top, palmar at the bottom. The left column shows the extracted ROI, the right column the extracted MC features

	Session 1			
	EER	FMR100	FMR1000	ZeroFMR
MC	**4.13 (±1.11)**	6.79	10.54	12.50
PC	10.71 (±1.72)	15.54	18.39	18.75
GF	28.08 (±2.50)	68.39	74.64	75.89
SIFT	10.63 (±1.72)	17.50	27.68	28.57
DTFPM	7.33 (±1.45)	13.04	16.61	17.32

	Session 2			
	EER	FMR100	FMR1000	ZeroFMR
MC	**5.69 (±1.77)**	9.44	23.06	23.06
PC	8.97 (±2.17)	13.61	16.94	16.94
GF	36.37 (±3.66)	81.39	85.83	85.83
SIFT	14.17 (±2.65)	31.11	37.22	37.22
DTFPM	12.00 (±2.47)	23.33	28.33	28.33

	Session 1 vs Session 2			
	EER	FMR100	FMR1000	ZeroFMR
MC	**24.30 (±2.49)**	56.13	69.87	69.87
PC	28.48 (±2.62)	54.13	68.00	68.00
GF	42.24 (±2.87)	96.67	99.20	99.20
SIFT	41.12 (±2.86)	88.80	94.40	94.40
DTFPM	30.22 (±2.67)	65.87	74.53	74.53

Table 2: Recognition performance results in terms of EER/FMR100/FMR1000/ZeroFMR for the dorsal sub-set for the single sessions and cross session

GF cannot compete with the other methods. The DET plots for session 1 and 2 are depicted in Figure 6.

Considering that the images have been acquired fully contactless in an nearly unconstrained environment, the recognition rate of the system for the single individual ses-

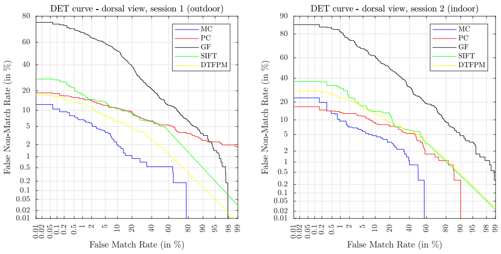


Figure 5: DET plot for session 1 (left) and session 2 (right) of the dorsal view

	Session 1			
	EER	FMR100	FMR1000	ZeroFMR
MC	**7.52 (±1.47)**	10.54	13.04	13.39
PC	13.88 (±1.93)	23.75	31.07	34.64
GF	32.52 (±2.61)	85.71	90.71	93.93
SIFT	11.90 (±1.80)	21.43	34.11	39.82
DTFPM	7.67 (±1.48)	12.14	16.79	21.96

	Session 2			
	EER	FMR100	FMR1000	ZeroFMR
MC	**7.78 (±2.04)**	15.28	22.78	22.78
PC	14.52 (±2.68)	21.94	24.17	24.17
GF	33.93 (±3.60)	82.22	89.17	89.17
SIFT	14.21 (±2.66)	30.28	43.61	43.61
DTFPM	12.14 (±2.49)	22.50	26.67	26.67

	Session 1 vs Session 2			
	EER	FMR100	FMR1000	ZeroFMR
MC	**27.73 (±2.60)**	56.00	65.47	65.47
PC	34.27 (±2.76)	62.80	75.33	75.33
GF	42.24 (±2.87)	98.53	99.87	99.87
SIFT	41.38 (±2.86)	86.00	95.87	95.87
DTFPM	34.07 (±2.76)	76.67	85.07	85.07

Table 3: Recognition performance results in terms of EER/FMR100/FMR1000/ZeroFMR for the palmar sub-set for the single sessions and cross session

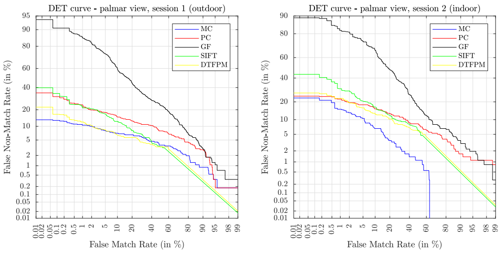


Figure 6: DET plot for session 1 (left) and session 2 (right) of the palmar view

sions is acceptably good. The inferior performance of session 2 (indoor with more controlled artificial ambient light) might be due to the proposed illumination add-on which does not provide enough NIR light to sufficiently highlight the "deeper" veins. The additional NIR light present in sunlight might help to render the veins more visible in the outdoor session and therefore increase its recognition performance. The palmar images are dominated by the creases of the hand. The veins on the palmar side are deeper inside the skin as on the dorsal side. Our illumination does not penetrate deep enough into the tissue. Therefore, the vein structure is only partially visible. This explains the performance decrease of the palmar sub-set compared to the dorsal one.

The significant performance drop of the inter-session comparison might result from the unconstrained environment. Vein structure based methods rely on the correlation of the images. During comparison we shifted the images in x- and y-direction and rotated them in order to maximise the correlation and correct small displacements. This corrections might have not been enough as they do not consider non-planar rotations (tilt). Considering our previous work [18], one could expect that key-point based algorithms, especially DTFPM, are better suited for an unconstrained acquisition environment. The performance of these methods needs to be further investigated.

5.4. Challenge-Response Evaluation Results

The challenge-response evaluation has been conducted on a total of 65 videos for 13 different users with random blinking sequences. For this purpose, the expected binary sequences determined by the random number generated for each video have been compared to the detected binary sequences, which have been extracted by the procedure described in Section 3.1, by means of Hamming Distance (HD). A sequence has been defined as a match (M), only with a HD equal to 0. Otherwise, the detected sequence has been defined as a non-match (NM). Table 4 shows the matching accuracy and mean hamming distance for non-matches for each user. A mean detection accuracy of 0.82 has been achieved for all CR-videos with a mean HD of 3.04 for failed detection attempts. Compared to a related biometric recognition and CR solution proposed in [20], where a detection accuracy of 0.40 has been achieved in one CR authentication attempt (80 out of 201), we obtain a competitive result.

The failed detection is mainly caused by synchronisation problems, e.g. for users 3 and 10. This synchronisation problems arise from two different factors: a software factor in form of the used camera library and a hardware problem with the timing of the LEDs. The camera library causes some delay with the video acquisition, which causes a desynchronisation of the blinking intervals, while the current implementation of the embedded LED control can cause inconsistent timer intervals with a deviation of up to 200 ms. Furthermore, due to the missing manual exposure setting in the camera library, some longer consecutive sequences of 0s or 1s are not interpolated correctly. All of these issues will be addressed in future versions of the mobile hand-vein scanner add-on, by choosing another library to access the camera and changing the embedded LED control to an interrupt based control.

User	ACC	Mean HD NM
1	0.60	2
2	1.00	-
3	0.80	4
4	0.80	2
5	0.60	2.50
6	0.20	2.75
7	1.00	-
8	0.80	2
9	1.00	-
10	0.80	6
11	1.00	-
12	1.00	-
13	1.00	-
Mean	0.82	3.04

Table 4: Detection accuracy (ACC) and mean Hamming Distance for non-matches (Mean HD NM) for CR sequence detection. If $ACC = 1$ there are no non-matches, so no distance can be calculated, thus there is $-$ in the Mean HD MM column.

6. Conclusion and Future Work

We proposed an illumination add-on for smartphones which turns a smartphone with an NIR-sensitive camera into a mobile hand vein scanner device. Using such a scanner, we established a publicly available data set acquired in two time-span separated and environmental different (indoor, outdoor) sessions and analysed the recognition performance of the new data set utilising some well-established vein recognition schemes. We further proposed a challenge-response protocol in order to prevent replay and presentation attacks and evaluated its applicability.

In our future work we will further develop our illumination add-on to enhance the acquisition quality. We will look into a multi-sample fusion of the different video frames captured from the hand-veins in order to improve the recognition performance. Moreover, we aim to evolve DTFPM as a hand vein recognition scheme which is tolerant against non-planar rotations. In addition, we will continue to develop our challenge-response protocol, improve the Android app and LED controller. Furthermore, we plan to utilise the smartphone's built-in sensors to deal with some of the imposed challenges caused by the unrestricted positioning of the phone relative to the hand. After all of the mentioned improvements have been implemented, we will further extend our mobile hand-vein data set by acquiring additional subjects.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 700259.

References

[1] Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S. E. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, L. Qin, F. Peng, L. B. Zhang, M. Long, S. Bhilare, V. Kanhangad, A. Costa-Pazo, E. Vazquez-Fernandez, D. Perez-Cabo, J. J. Moreira-Perez, D. Gonzalez-Jimenez, A. Mohammadi, S. Bhattacharjee, S. Marcel, S. Volkova, Y. Tang, N. Abe, L. Li, X. Feng, Z. Xia, X. Jiang, S. Liu, R. Shao, P. C. Yuen, W. R. Almeida, F. Andalo, R. Padilha, G. Bertocco, W. Dias, J. Wainer, R. Torres, A. Rocha, M. A. Angeloni, G. Folego, A. Godoy, and A. Hadid. A competition on generalized software-based face presentation attack detection in mobile scenarios. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 688–696, Oct 2017.

[2] J. H. Choi, W. Song, T. Kim, S.-R. Lee, and H. C. Kim. Finger vein extraction using gradient normalization and principal curvature. *Proc.SPIE*, 7251:7251 – 7251 – 9, 2009.

[3] P. C. Eng and M. Khalil-Hani. Fpga-based embedded hand vein biometric authentication system. In *TENCON 2009 - 2009 IEEE Region 10 Conference*, pages 1–5, Jan 2009.

[4] R. R. Fletcher, V. Raghavan, R. Zha, M. Haverkamp, and P. L. Hibberd. Development of mobile-based hand vein biometrics for global health patient identification. In *Global Humanitarian Technology Conference (GHTC), 2014 IEEE*, pages 541–547. IEEE, 2014.

[5] V. Kanhangad, A. Kumar, and D. Zhang. Contactless and pose invariant biometric identification using hand surface. *IEEE transactions on image processing*, 20(5):1415–1424, 2011.

[6] C. Kauba, B. Prommegger, and A. Uhl. Focussing the beam - a new laser illumination based data set providing insights to finger-vein recognition. In *Proceedings of the IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS2018)*, pages 1–9, Los Angeles, California, USA, 2018.

[7] C. Kauba, J. Reissig, and A. Uhl. Pre-processing cascades and fusion in finger vein recognition. In *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG'14)*, Darmstadt, Germany, Sept. 2014.

[8] M. Khalil-Hani and P. C. Eng. Fpga-based embedded system implementation of finger vein biometrics. In *2010 IEEE Symposium on Industrial Electronics and Applications (ISIEA)*, pages 700–705, Oct 2010.

[9] M. Khalil-Hani and P. C. Eng. Personal verification using finger vein biometrics in fpga-based system-on-chip. In *2011 7th International Conference on Electrical and Electronics Engineering (ELECO)*, pages II–171–II–176, Dec 2011.

[10] P. Korshunov, S. Marcel, H. Muckenhirn, A. Gonçalves, A. Mello, R. Violato, F. Simões, M. Uliani Neto, M. de Assis Angeloni, J. A. Stuchi, et al. Overview of btas 2016 speaker anti-spoofing competition. Technical report, Idiap, 2016.

[11] A. Kumar and Y. Zhou. Human identification using finger images. *Image Processing, IEEE Transactions on*, 21(4):2228–2244, 2012.

[12] H. C. Lee, K. R. Park, B. J. Kang, and S. J. Park. A new mobile multimodal biometric device integrating finger vein and fingerprint recognition. In *Ubiquitous Information Technologies & Applications, 2009. ICUT'09. Proceedings of the 4th International Conference on*, pages 1–4. IEEE, 2009.

[13] Z. Liu and S. Song. An embedded real-time finger-vein recognition system for mobile devices. *IEEE Transactions on consumer Electronics*, 58(2), 2012.

[14] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision (CVPR'99)*, volume 2, pages 1150 – 1157. IEEE, 1999.

[15] Y. Matsuda, N. Miura, A. Nagasaka, H. Kiyomizu, and T. Miyatake. Finger-vein authentication based on deformation-tolerant feature-point matching. *Machine Vision and Applications*, 27(2):237–250, 2016.

[16] N. Miura, A. Nagasaka, and T. Miyatake. Extraction of finger-vein patterns using maximum curvature points in image profiles. *IEICE transactions on information and systems*, 90(8):1185–1194, 2007.

[17] A. Morales, M. A. Ferrer, and A. Kumar. Towards contactless palmprint authentication. *IET computer vision*, 5(6):407–416, 2011.

[18] B. Prommegger, C. Kauba, and A. Uhl. Longitudinal finger rotation - problems and effects in finger-vein recognition. In *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG'18)*, Darmstadt, Germany, 2018.

[19] A. Sierro, P. Ferrez, and P. Roduit. Contact-less palm/finger vein biometrics. In *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–12, Sept 2015.

[20] C. Stein, V. Bouatou, and C. Busch. Video-based fingerphoto recognition with anti-spoofing techniques with smartphone cameras. In *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, pages 1–12, Sept 2013.

[21] P. Tome and S. Marcel. On the vulnerability of palm vein recognition to spoofing attacks. In *The 8th IAPR International Conference on Biometrics (ICB)*, May 2015.

[22] University of Reading. PROTECT Multimodal DB Dataset, June 2017. Available by request at http://projectprotect.eu/dataset/.

[23] J. Zhang and J. Yang. Finger-vein image enhancement based on combination of gray-level grouping and circular gabor filter. In *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*, pages 1–4. IEEE, 2009.

[24] J. Zhao, H. Tian, W. Xu, and X. Li. A new approach to hand vein image enhancement. In *Intelligent Computation Technology and Automation, 2009. ICICTA'09. Second International Conference on*, volume 1, pages 499–501. IEEE, 2009.

[25] K. Zuiderveld. Contrast limited adaptive histogram equalization. In P. S. Heckbert, editor, *Graphics Gems IV*, pages 474–485. Morgan Kaufmann, 1994.

Mobile Face Recognition Systems: Exploring Presentation Attack Vulnerability and Usability

Heinz Hofbauer
University of Salzburg
hofbauer@cs.sbg.ac.at

Luca Debiasi
University of Salzburg
ldebiasi@cs.sbg.ac.at

Andreas Uhl
University of Salzburg
uhl@cs.sbg.ac.at

Abstract

We have evaluated face recognition software to be used with hand held devices (smartphones). While we can not go into specifics of the systems under test (due to NDAs), we can present the results of our evaluation of liveness detection (or presentation attack detection), matching performance, and success with different complexity levels of attacks. We will contrast the robustness against presentation attacks with the systems usability during regular use, and highlight where currently state of commercial of the shelf systems (COTS) stand in that regard. We will look at the results specifically under the tradeoff between acceptance, linked with usability, and security, which usually negatively impacts usability.

1. Introduction

We were tasked by a company with evaluating the usability and security of face recognition systems which work by recording a selfie (self-portrait) on a smartphone. The matching was done on the server side, but liveness detection was done on the smartphone. The company ran the servers, provided the hardware and software. The whole project was on a rather tight time schedule (due to license lease time), so we could only conduct a limited number of experiments with a limited number of people. Nonetheless, the results were rather interesting and we wanted to share them.

That said, this is not a very technical paper. It is more a recording of our experience with the software/devices/processes. The main incentive to share this information is to showcase certain problems which do not happen in a typical “lab setup”. Shortcomings in algorithms or implementation can be detrimental to the adoption by industry or acceptance by users and it can occasionally lead to interesting research questions too. In this paper we will present our experiments and findings and comment on how research might help.

978-1-7281-3640-0/19/$31.00 ©2019 IEEE

We fill focus more on the what is of interest to us as researchers and less on implementation details, except where the used protocol might impact the research side. That said, we would like to point out that software implementations, even only research software for reproducible research, should be built with corner cases in mind to allow for testing on more difficult test sets*.

Limited Tests: Due to time constraints, only a short license lease time during which to test the systems was granted, we could only afford a very limited number of tests. Specifically, most test were only performed by a single user. The number of attempts was also rather low, usually 10 to 20 repeats per test. Yet, even with such a limited number of tests we could find counterevidence regarding the security of the systems.

The goal in all these tests is to have a method to unlock the device or otherwise verify the user of the device when the user cooperates. What is important to companies is that this process is secure on the one hand, but also fast and annoyance free for the user. If this latter part is not given, an adoption of the system by users is less likely.

As such we will look at the security of the two systems under test, PassiveSys and ActiveSys, with the goal of unlocking the device with minimal fuss on the part of the user.

The paper is structured as follows, Section 2 gives an overview of presentation attacks and their detection as it relates to the matter at hand. Section 3 establishes a baseline when genuine traits are presented to the systems under test. Section 4 will attack the test with replay type attacks and Section 5 will use more sophisticated replicas of the biometric traits to circumvent the system. Finally, Section 6 will summarize our findings and conclude the paper.

*While we will delve no deeper into this we just would like to note that we managed to crash the server because of the floral pattern design on a user’s shirt worn during testing

2. Related Work

Smartphones are ubiquitous and so is the widespread adoption of biometric traits to unlock the device by verifying the identify of the user. In recent years, a certain trend from using fingerprints towards face detection can be observed. This trend has renewed the interest in attacks, and the prevention thereof, against such biometric systems.

A specific attack is the presentation, also known as direct or spoofing, attack. It can be separated into two categories [1]: (1) active imposter presentation attacks, where the attacker tries to claim a foreign identity; and (2) concealer presentation attacks, where an attacker tries to not be recognized by a system. Presentation attacks can be used against identification as well as verification modes.

Presentation attacks (PA) can also be differentiated by the source of the presentations attack instrument (PAI): (1) artificial, which is a non-human material sourced from humans, e.g., masks, printouts, images; (2) human traits, parts of dead bodies, modified faces, forced presentation by unconscious persons and so on.

To prevent such attacks, a presentation attack detection (PAD) system, also referred to as liveness detection, is employed. The primary focus of research is artificial presentation but, as is evident in the term liveness detection, overlaps with parts of the human trait PAI categorization.

There are different kinds of (face detection) PADs, some are hardware reliant while others are not, some use still images and others video. The number of different PAD methods is long, thus we will only give a brief list of methods without going into them too much: blink detection ([2, 3]), challenge response ([4–6]), texture based ([7–9]), dynamic texture based (video) ([10, 11]) or movement based ([12–14]). For more details, the reader is referred to the respective papers.

The target application of our tests was to unlock the device with the presented biometric trait (face). The operation mode, in terms of biometry, is always verification since the identify is implied (the owner of the cell phone). Presentation attacks also try to unlock the device and are consequently also done in verification mode. The presentation attack instrument is artificial only. While there are more types of PAIs, and a lot of further differentiation by subtype, we only gave related literature to the modes suspected to be employed in the devices we test. Specifically, ActiveSys certainly uses blink detection and challenge response methods. PassiveSys's modes are all passive, i.e., no cue based user interaction is required, using image, and we strongly suspect video, and thus has to rely on texture based image and video as well as movement features for PAD.

Please note: In the following sections we will present tables with results. These results are in the form of success rate of the liveness detection (LD) and the match rate (MR), which relate to the reporting as specified in ISO/IEC 30107-3 [15] as follows: In case the presented trait was genuine, the bona fide presentation classification error (BPCER) can be calculated as $\mathrm{BPCER} := 1 - \mathrm{LD}$. In case of presentation attacks the attack presentation classification error rate (APCER) can be calculated as $\mathrm{APCER} := 1 - \mathrm{LD}$. Likewise, the false non match rate (FNMR) for genuine presentation is $\mathrm{FNMR} := 1 - \mathrm{MR}$ and the impostor attack presentation match rate (IAPMR) for presentation attacks is $\mathrm{IAPMR} := \mathrm{MR}$.

3. Usability

For usability, we look at the basic modes provided by the software. With these modes we get a baseline for further tests and presentation attacks. We evaluated two software systems, denoted PassiveSys and ActiveSys, both have a separate step for detecting liveness and matching the probe and gallery image.

PassiveSys could operate with five different modes, which only impact liveness detection. No further detail on what is different was provided to us, but on-screen notes gave clues on what is required for the liveness detection. Modes are: *video*, unclear conditions but seems to take a video; *lessvid*, seems to be a less stringent version of *video*; *image*, simply takes a picture. One mode was not used because we could never pass liveness detection. Another mode was designed to use the rear camera and an operator to identify a second person, this was not used because the goal is to unlock the device (single user operation).

ActiveSys allows four liveness detection modes: None; *blink*, user has to keep still and blink on cue; *arrow*, requires turning the head to steer an arrow along a line to a target, when the arrow and target align the user has to blink; *blink+arrow*, a combination of both modes. We will not give separate results for None and the *blink+arrow* combination since the modes are simply executed one after the other.

3.1. Usability and Baseline

To get a baseline for the systems, we created two test sets, one where the gallery images is from a user with glasses and one where the user does not wear glasses.

The results are given in Table 1, split for system and liveness detection type. It can be seen that the presence of glasses in the image increases the error rate of the liveness detection. It is also interesting to see that the matching always worked when liveness detection was passed. However, even for probe images without glasses certain modes did reject a lot of attempts, *video* overall rejected almost 72% of all attempts. Also interesting is that *arrow* seems to reject less attempts than *blink*, even though the task is more complicated. The modes of PassiveSys on the other

Table 1: Baseline for the ActiveSys and PassiveSys. Results are split between liveness detection test (LD) and verification results (Match). The presence of glasses in the probe (Pr.) and gallery (Gal.) images is given as well.

(a) Baseline for PassiveSys for modes *video*, *lessvid*, *image*.

video

Pr.	Gal.	LD	Match
yes	yes	0/20	0/20
no	no	13/20	13/20
no	yes	10/20	10/20
yes	no	0/20	0/20

lessvid

Pr.	Gal.	LD	Match
yes	yes	4/20	4/20
no	no	18/20	18/20
no	yes	12/20	12/20
yes	no	9/20	9/20

image

Pr.	Gal.	LD	Match
yes	yes	6/20	6/20
no	no	20/20	20/20
no	yes	18/20	18/20
yes	no	19/20	19/20

(b) Baseline for ActiveSys for modes *blink* and *arrow*.

blink

Pr.	Gal.	LD	Match
yes	yes	20/20	20/20
no	no	16/20	16/20
no	yes	12/20	12/20
yes	no	12/20	12/20

arrow

Pr.	Gal.	LD	Match
yes	yes	18/20	18/20
no	no	20/20	20/20
no	yes	20/20	20/20
yes	no	20/20	20/20

Table 2: Performance during bright sunlight outdoors. Results are split between liveness detection test (LD) and verification results (Match). Facing was either towards the sun or away from the sun.

(a) PassiveSys split for modes.

video and *lessvid*

Facing	LD	Match
towards	0/20	0/20
away	0/20	0/20

image

Facing	LD	Match
towards	10/20	10/20
away	20/20	20/20

(b) ActiveSys split for modes.

blink

Facing	LD	Match
towards	14/20	14/20
away	9/20	9/20

arrow

Facing	LD	Match
towards	10/20	10/20
away	19/20	19/20

Table 3: Liveness Detection under studio light for different light positions (Dir.) and intensities. Light was diffused or un-diffused as a spot light. Entries are the number of success based on 10 attempts per setting.

			LD under Intensities					
			spot			diffuse		
System	Mode	Dir.	1.0	3.0	6.0	1.0	3.0	6.0
PassiveSys	*video*	front	0	0	0	0	0	0
PassiveSys	*lessvid*	front	4	2	3	10	5	5
PassiveSys	*image*	front	8	8	9	10	9	10
ActiveSys	*blink*	front	8	8	6	2	5	3
ActiveSys	*arrow*	front	9	9	10	10	7	8
PassiveSys	*video*	side	0	0	0	0	0	0
PassiveSys	*lessvid*	side	3	4	2	7	5	0
PassiveSys	*image*	side	9	8	9	10	7	8
ActiveSys	*blink*	side	3	4	4	4	4	6
ActiveSys	*arrow*	side	3	5	1	10	5	3
PassiveSys	*video*	back	0	0	0	0	0	0
PassiveSys	*lessvid*	back	5	3	1	4	3	1
PassiveSys	*image*	back	5	9	6	2	0	1
ActiveSys	*blink*	back	6	7	5	7	5	3
ActiveSys	*arrow*	back	9	10	10	10	9	10

hand behave as expected, the more complicated method reject more attempts, i.e., video based reject more than image base liveness detection modes.

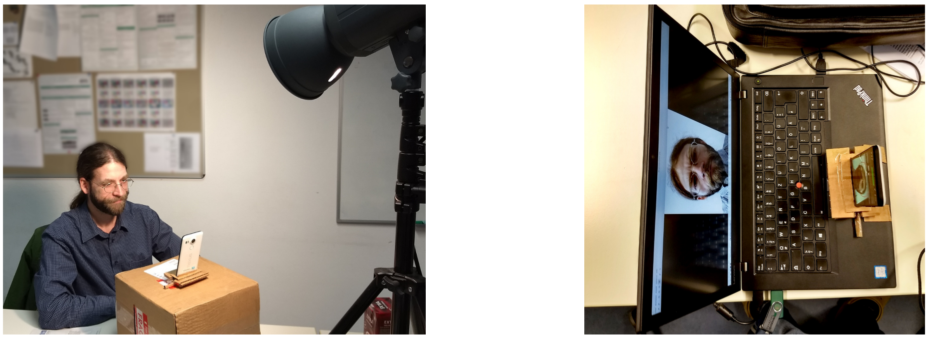
What also resulted from these experiments, which is not reflected in the table, is the insight that failure of longer modes, like the *arrow* or *blink+arrow* modes for ActiveSys which took several seconds per attempt, became frustrating very fast.

3.2. Usability and Baseline Outdoors

We suspected that the failure to detect images with glasses as alive was due to reflection of light on the glasses. The results in Table 1 were obtained from experiments in a well lit room. To further test the impact of light on the liveness detection and to expand the baseline to the outdoors, we performed another test in natural sunlight, during a bright day.

This experiment was conducted without glasses and the results are given in Table 2. The clear impact of lighting conditions on the liveness detection is quite drastic, *video* and *lessvid* failed to detect anything as alive and *image*, *blink* and *arrow* all had reduced number successful attempts. However, it should also be noted that the actual verification always worked when the liveness detection was passed. This might be a benefit of the aggressive screening during liveness detection, which is not necessarily a bad thing since early failure is less costly in terms of time to failure.

To get a more reproducible, and finer grained, version of the light test we set up a dimmed room with a studio light (Helios 300p) shining at the user from the front, side or back at a distance of roughly 1m. The light levels were adjustable and were set to 1, 3 and 6 (from a maximum setting of 6) and we investigated spot and diffuse (diffused with bleached



(a) Evaluation of the impact of frontal studio light.

(b) Replay attacks for still images, same setup was also used for degraded images.

Figure 1: Different test setups.

80g/m^2 paper) light to simulate a clear or cloudy day, one such experiment is depicted in Figure 1a. The results are given in Table 3, verification is not given separately since every time liveness was detected the user was also correctly verified. Results given are the successful unlock attempts from a set of 10 attempts per parameter set.

The results from the controlled tests show quite nicely the influence of light on the different modes. All modes are affected to some degree and for the most part how they are affected makes sense, e.g., higher effect the stronger the light is, spot light has a higher effect than diffuse light and so on. The one difference is the direction, frontal light illuminates the subject unlocking the device so has the least influence, but the higher reduction of side illumination over backlight is somewhat surprising. The sidelight usually results in a very uneven illumination, one face side in shadow the other illuminated. Backlight should mess up the exposure settings of the camera and leave the whole face in shadow. The expectation therefore would be that the backlight exhibits worse performance than sidelight, which is not backed by experimental results.

3.3. Discussion

Time to failure and repeats can heavily impact the user experience. The overall time taken to unlock has to be acceptable to the user. Failures do not matter so much, so if failure and retry is fast and painless then the resulting user experience can still be good. However, if a long process fails and has to be retried, user satisfaction quickly fades. On a related note users try to help the system by doing the "right" thing to speed up the process. We can use this by making explicit what is required rather than letting the user guess. The user is a willing participant and will try to help as much as possible to speed up the unlocking process.

Light and the outdoors environment. The impact of directional light on the liveness detection system is quite drastic and will make many of the modes under test unfeasible in practice. And while the matching worked well for all cases it is not clear if this is due to the aggressive liveness screening or robust matching algorithms.

4. Presentation Attack: Replay Attacks

The next logical step to test the security of the system was to perform a replay attack. That is, record an image or video and present that to the device instead of the genuine face. In a perfect world the liveness detection should reject every attempt.

To reduce the amount of data to display in tables, the *video* mode will no longer be used. Given its problems of rejecting images with glasses and strong light, it will likely never be used in practice either.

For the simple replay attack, we used a printed version of the image, the image displayed on a computer screen and a short video also displayed on the screen. The latter was used since both modes from ActiveSys require at least blinking and a bit of interaction in the case of *arrow*, simulated by turning the smartphone. The setup for the static image replay attacks and test of degradation types (see below) is shown in Figure 1b.

The results from this test can be seen in Table 4, again only liveness detection is given since verification was always successful when liveness was detected.

It is interesting to compare these results to the lighting results in Table 2. The same stringency which allows the detection of replay attacks adversely affects the usability in environments with bright lights. Overall, the expected result is present, higher quality/effort reproductions have a higher success chance, i.e., video is better than screen is better than

Table 4: Result of a replay attack. The number of successful attacks out of 20 attempts is given.

		Successes with Replay		
System	Mode	print	screen	video
PassiveSys	*lessvid*	0	1	0
PassiveSys	*image*	12	17	20
ActiveSys	*blink*	—	—	0
ActiveSys	*arrow*	—	—	5

Table 5: Result of a controlled degraded image replay attack. The number of successful attacks out of 10 attempts is given.

			Strength of Degradation		
System	Mode	Degradation	low	medium	high
PassiveSys	*image*	Noise	10	10	10
PassiveSys	*image*	Blur	10	10	6
PassiveSys	*image*	Resolution	10	10	6/0*

*liveness was detected 6 times, but verification was passed 0 times

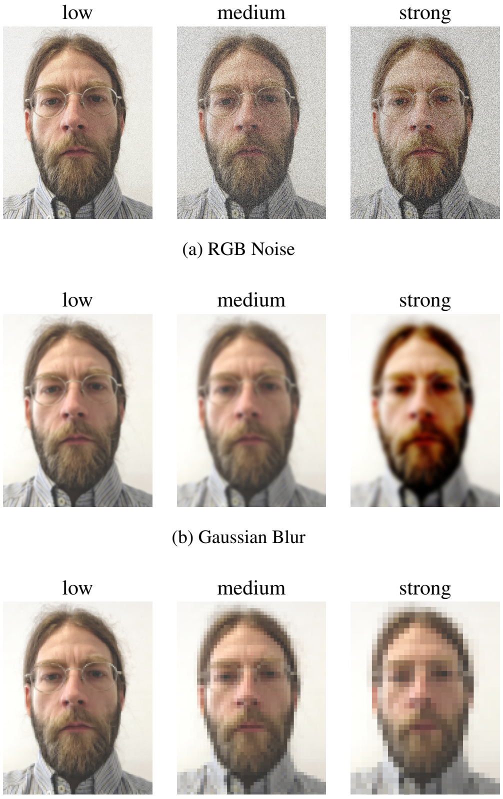
print. And again the *arrow* mode is easier to pass than the *blink* mode, even though more 'user' interaction is required.

Assuming that usability is a prime factor for a widespread adoption of such unlock systems, we will take a closer look at just how bad a recording still allows an unlock. From a practical perspective we will only look at the *image* mode. While *lessvid* would also be an interesting candidate, the mounting of such a replay attack is harder since a video has to be acquired, while *image* only requires a still image, i.e., a simple photograph. To simulate bad recording conditions we will add noise, blur the image and pixelate it to simulate a low resolution. The results are given in Table 5 for the *image* mode, Figure 2 illustrates the range of noise, blur and pixelation applied.

The clear result of these tests is that even a strongly degraded version of the image can penetrate the liveness detection of the *image* mode.

4.1. Discussion

An interesting tradeoff between usability and security can be observed in these experiments. Since usability is paramount for applicability, the security has the be reduced somewhat. However, this can be counteracted by user participation in activity assisted unlock modes like *arrow*. The drawback of such methods is that they take longer and require more attention from the user making a failure to unlock more annoying. This annoyance could hinder adoption of such schemes, which in turn would require a reduction in security and thus brings us full circle again. There is clearly a need for fast and reliable liveness detection methods.



(a) RGB Noise

(b) Gaussian Blur

(c) Resolution

Figure 2: Illustration of Degradation types and strength for the replay attacks in Table 5.

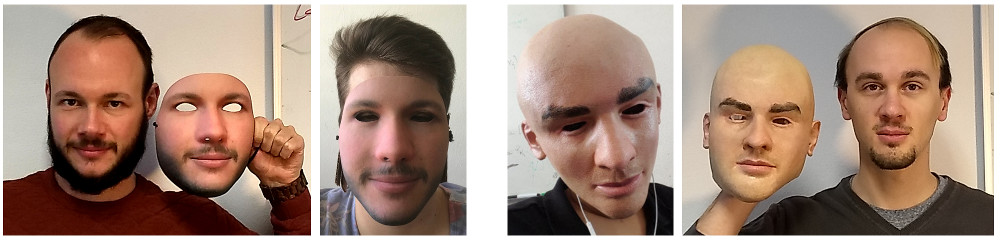
5. Presentation Attack: Masks

For these presentation attacks, we used a mask or mask-like presentation of the stolen biometric trait, created via photographs of the target's face. This was done to increase the chance of breaking interactive systems and give the impression of depth a 2D image might not convey.

We used two attack types: (1) a handcrafted 3D latex based mask by CREA FX*; and (2) a 3D-printed hard resin composite mask by ThatsMyFace†. Figure 3 show examples of the different masks.

Since the masks allow some interaction, we will again use both modes from ActiveSys as well as *image* and *lessvid*

*https://www.creafx.com/en/
†http://thatsmyface.com/custom-wearable-masks/


Resin composite

Latex

Figure 3: Masks used for presentation attacks. The sources of the biometric traits hold their replicas and imposter wearing the replicas during an attack attempt.

Table 6: Presentation attack results for both mask types and the given modes and systems.

		Latex Mask		Resin Mask	
System	Mode	LD	M	LD	M
PassiveSys	*lessvid*	5	0	10	0
PassiveSys	*image*	10	0	20	0
ActiveSys	*blink*	0	0	10	0
ActiveSys	*arrow*	0	0	16	0

from PassiveSys. The results are given in Table 6 out of 20 attempts.

The obtained results are very interesting, especially in comparison to prior experiments. Where until now the liveness detection was relatively stringent and the verification always worked when the liveness detection was passed, the table has turned here. The liveness detection, which really should catch these cases fails and lets them pass, while the verification rejects the masks.

There is also quite the difference in mask quality, while the latex mask was handcrafted, took about three times as long to acquire and was four times as expensive as the 3d-printed resin mask, it performed worse.

5.1. Discussion

What is interesting here is that the relatively high cost only marginally increases the success rate. To illustrate this, let us have a look at the threat level model laid out in [16], briefly given in table 7.

Table 7: Spoof presentation attacks separated by levels based on time, expertise, and equipment.

Threat	Level A	Level B	Level C
Time	short	>3 days	>10 days
Expertise	anyone	practice needed	extensive skill required
Equipment	readily available	requires planning	specialized
Biometric source	readily available	difficult to obtain	difficult to obtain
Example	paper print of image	paper mask or video	3D face reconstruction

Table 8: Comparison of threat level and success rate per mode and system. Usability, the chance of unlock by a genuine user is a combination of results from Table 1 and 2.

System	Mode	Attack	Threat	Success Rate	Usability
PassiveSys	*lessvid*	Image	Level A	5%	44.2%
PassiveSys	*lessvid*	Video	Level B	0%	
PassiveSys	*lessvid*	Mask	Level C	0%	
PassiveSys	*image*	Image	Level A	85%	77.5%
PassiveSys	*image*	Video	Level B	100%	
PassiveSys	*image*	Mask	Level C	0%	
ActiveSys	*blink*	Video	Level B	0%	69.2%
ActiveSys	*blink*	Mask	Level C	0%	
ActiveSys	*arrow*	Video	Level B	25%	89.2%
ActiveSys	*arrow*	Mask	Level C	0%	

Table 8 compares this to the results from our test, where success rate is the percentage of presentation which passed the liveness detection and verification. Usability is the chance of unlock by a genuine user under different conditions. What is most interesting is that Level C attacks, which are much more expensive and have a much higher preparation time, do not improve in success rate over Level B and Level A attacks.

From this table it also becomes clear that the PassiveSys system is basically unusable, either the usability of a given mode is low (*lessvid*) or the success rate of attack is high (*image*). The ActiveSys system is far better designed in this regard. A tradeoff between reduced usability and higher security (*blink*) and higher usability at the cost of a potential Level B attack (*arrow*) can be observed.

6. Conclusion

What we have seen is that biometric verification for the systems under test seems to work well. However, it is unclear if this is in part due to the strict liveness detection. While this may seem an odd differentiation, we have also seen that a strict liveness detection can reduce usability. At times this reduction can be quite drastic and based on plain and simple factors, like wearing glasses or trying to unlock the device during a bright day. As such, a step towards a higher usability and consequently user satisfaction and acceptance, would be to tweak the liveness detection to be less strict in such cases. However, if this has a negative effect on the matching performance, nothing is gained in terms of usability at the cost of security.

That said, the liveness detection of both tested systems does a relatively good job of screening attacks. Again, this success in screening attacks is at the cost of usability. While this tradeoff is fine in theory, the practical impact is quite high,i.e., PassiveSys reduced the chance of success for genuine presentations to less than 50% and could still be successfully attacked. While ActiveSys fared better, it also had to reduce the usability to around 70% to prevent attacks. There is clearly ample room for improvement.

Regarding the attacks, it was interesting to see that the most expensive and time consuming attacks, specially created facial masks, fared worse than relatively simple printed image or video presentation attacks.

Acknowledgements

This work has been partially supported by Veridos Inc.

References

[1] R. Raghavendra and C. Busch, “Presentation attack detection methods for face recognition systems: A comprehensive survey”, *ACM Computing Surveys*, vol. 50, no. 1, 2017.

[2] P. Gang, S. Lin, W. Zhaohui, and L. Shihong, “Eyeblink-based anti-spoofing in face recognition from a generic webcamera”, ser. Proceedings for IEEE 11th International Conference on Computer Visio, 2007.

[3] M. Chrzan, “Liveness detection for face recognition”, Master’s thesis, Masaryk University, Factory of Informatics, 2014.

[4] K. Kollreider, H. Fronthaler, and J. Bigun, “Verifying liveness by multiple experts in face biometrics”, ser. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008.

[5] A. Ali, F. Deravi, and S. Hoque, “Directional sensitivity of gaze-collinearity features in liveness detection”, in *4th International Conference on Emerging Security Technologies*, 2013.

[6] D. Smith, A. Wiliem, and B. Lovell., “Face recognition on consumer devices: Reflections on replay attacks”, *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, 2015.

[7] N. Kose and J. Dugelay, “Classification of captured and recaptured images to detect photograph spoofing”, in *International Conference on Informatics, Electronics Vision (ICIEV’12)*, 2012.

[8] J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection with component dependent descriptor”, in *International Conference on Biometrics (ICB’13)*, 2013.

[9] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face spoofing detection using colour texture analysis”, *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, 2016.

[10] I. Chingovska, A. André, and S. Marcel, “On the effectiveness of local binary patterns in face anti-spoofing”, in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG ’12)*, 2012.

[11] A. A. Tiago de Freitas Pereira, J. M. D. Martino, and S. Marcel, “LBP-TOP based countermeasure against face spoofing attacks”, in *Computer Vision - ACCV 2012 Workshops*, ser. Lecture Notes in Computer Science, vol. 7728, 2012.

[12] M. D. Marsico, M. Nappi, D. Riccio, and J. Dugelay, “Moving face spoofing detection via 3d projective invariants”, in *5th IAPR International Conference on Biometrics (ICB’12)*, 2012.

[13] A. Anjos, M. M. Chakka, and S. Marcel, “Motion-based counter-measures to photo attacks in face recognition”, *IET Biometrics*, 3 2014.

[14] A. Pinto, W. Schwartz, H. Pedrini, and A. Rocha, “Using visual rhythms for detecting video-based facial spoof attacks”, *EEE Transactions on Information Forensics and Security*, vol. 10, no. 5, 2015.

[15] ISO/IEC 30107-3, *Information technology - biometric presentation attack detection - part 3: Testing and reporting*, Sep. 2017.

[16] S. Schuckers, “Presentations and attacks, and spoofs, oh my”, *Image and Vision Computing*, vol. 55, pp. 26–30, 2016, Recognizing future hot topics and hard problems in biometrics research. DOI: 10.1016/j.imavis.2016.03.016.

4. Conclusion

The main contributions of the work conducted during my thesis focus on three main areas: biometric sensor forensics, face morph detection and mobile applications.

Biometric sensors acquiring image data are deployed in many biometric systems in order to capture biometric samples. Identifying the source sensor from the acquired image samples can offer advantages to the overall system in terms of improved security and interoperability. Therefore, we investigated the differentiability of iris and fingerprint sensors, which is essential for reliable source sensor authentication. Previous work found strong variations in the differentiability of sensors in certain datasets. First, we evaluated these challenging datasets by applying and extending established clustering techniques using the photo response non-uniformity (PRNU) to detect the number of sensors used to acquire the data. We assumed that the variations might originate from multiple sensors of the same model being used during data acquisition. In addition, a broad range of PRNU enhancement techniques and their effectiveness has been assessed in the context of biometric data. We identified that in general the cohesion and separation of the obtained clusters is improved, but in a highly situational manner. Continuing these efforts, we investigated a real world criminal case dataset containing still images found on a suspect's computer with the goal to cluster the images by source camera. In this context, we performed an evaluation of various cluster validity indices (CVIs) and source sensor clustering techniques, revealing appropriate clustering algorithms and CVIs for such a scenario. We also proposed an alternative image origin identification technique for iris images and compared it to the classical PRNU-based approach, where we identified the strengths and weaknesses of both approaches in different applications.

The second area focused on the detection of morphing attacks on face recognition systems. We proposed a novel PRNU-based morph detection algorithm that analyses alterations of the PRNU caused by the morphing process. We investigated different spectral and spatial features and extended the approach to detect image region specific variations. The proposed morph detector has proven to be robust against a wide range of image post-processings, morphing techniques (landmark and neural network based) as well its generalisability regarding the camera used to acquire the images. The proposed detector was able to significantly outperform other state-of-the-art morph detectors in scenarios where the image source or morphing technique is unknown.

The remainder of my thesis focused on mobile devices (smartphones) and covered diverse applications in this context. We evaluated multiple face recognition solutions for mobile devices with focus on their usability and security in terms of matching performance and their ability to detect presentation attacks with different complexity. We observed that the systems are able to achieve high levels of security, but only by compromising the usability. We also explored an alternative to classical biometric authentication such as fingerprints and face by designing and constructing a prototype NIR-illumination add-on for smartphones which enables the acquisition of vascular patterns. We acquired a hand-vein dataset and evaluated the recognition performance with well established vein recognition algorithms. In addition, we also proposed a challenge response protocol to ensure the authenticity of the acquired data based on variation of the NIR illumination. Finally, we explored an alternative and innovative application for mobile devices: counterfeit drug detection. We developed an authentication system that analyses the intrinsic texture features of the drug's packaging materials (packaging and blister) without the use of any external markers. We observed from our experiments that these intrinsic features are highly discriminative and constant across multiple packages of the same drug manufacturer. Therefore, we demonstrated the feasibility of using smartphones for this application and similar ones.

4.1. Issues and open challenges

While analysing biometric sensors based on their PRNU, iris and fingerprint sensors in particular featured strong variations in regard to their differentiability. Applying different PRNU enhancement techniques helped in attenuating several artefacts in the extracted PRNU, which lead to an increased differentiability improving the device identification and clustering performance. However, some of the investigated datasets still showed unclear results. We assume that they could either originate in non-unique artefacts (NUAs) within the extracted PRNU or might also be caused by a misalignment of the PRNU signatures. These two assumptions are made on the basis of investigating existing datasets with unknown acquisition conditions and experiences gathered from using various biometric sensors. Analysing NUAs might be very challenging in this scenario because of the highly correlated image content of iris and fingerprint samples. Misalignment issues might be caused by the way some sensors process the acquired image data, e.g. the sensor could track an eye across the image area and only crop a specific region, which can therefore hardly be corrected or attenuated. This highlights the need for a dataset with images from a large number of biometric sensors, as already exist for benchmarking forensic schemes in the context of consumer cameras [35] and smartphone cameras [66], with known ground truth and controlled acquisition conditions, in order to address the open issue of differentiability of biometric sensors. Furthermore, images with uncorrelated content, such as acquired in [23] would enable an examination of potential NUAs and thus help in shedding light on these issues. Unfortunately, existing biometric datasets have been acquired with focus on investigating the biometric samples and not the sensors. In contrast, the differentiability of consumer cameras based on their PRNU has been extensively demonstrated and many device identification and clustering techniques have been proposed in literature. However, real world applications as performed in [19] illustrate the limitations of these approaches. Datasets in literature usually contain images from different cameras, which are evenly distributed among them and have been acquired under controlled conditions using the base ISO sensitivity of the cameras. In real world datasets, though, acquisition conditions are usually unknown. These might include post-processings, such as cropping, scaling, rotation, contrast enhancement and other transformations as well as a wide range of different ISO sensitivities. To the best of the authors knowledge, the robustness of the PRNU has not been studied extensively in this regard, but clearly needs further investigation.

The proposed PRNU-based face morph detection system is shown to outperform other detectors in terms of generalisablity and robustness to different attacks. These two aspects are often neglected in current literature. As discussed in [59], the robustness of the system in regard to intentional and unintentional attacks on the PRNU has not yet been investigated. Another issue in the field is the lack of a publicly available benchmarking dataset and evaluation protocols to facilitate comparing different detection approaches and their performance. Some efforts are already being made in this regard by NIST with the FRVT MORPH project[1]. It also includes print-scan morphing attacks, which are inherently more challenging to detect compared to purely digital morphs due to additional artefacts being introduced by the printing and scanning process. The ability to detect print-scan morphing attacks plays a vital role in the detection

[1] https://pages.nist.gov/frvt/html/frvt_morph.html

scheme's real-world applicability for ePassports, since morphed face images might not only be included into the chip portion of the passport, but the photo printed on the passport might also be attacked with morphed images.

As previously mentioned, smartphones enable a wide range of applications. Because of their important role in authenticating ones identity, the employed biometric systems, i.e. mainly fingerprint and face recognition, need to withstand different types of attacks. We demonstrated that many mobile face recognition systems are vulnerable to presentation attacks in [41], despite claiming robustness against such attacks. Clearly, more sophisticated presentation attack detection approaches need to be developed without compromising the usability of the overall system. In order to achieve these goals, including dedicated hardware into the mobile devices might be inevitable. Alternatively, other biometric traits, such as vascular patterns, need to be investigated in the context of mobile devices. However, the capturing of alternative biometric traits might require additional hardware to be integrated as well. Nonetheless, mobile devices in their current state already offer and endless variety of applications and have become an integral part of this world, which is going to increase even further.

Bibliography

[1] A. Agarwal, R. Singh, M. Vatsa, and A. Noore. Swapped! digital face presentation attack detection via weighted local magnitude pattern. In *2017 IEEE International Joint Conference on Biometrics (IJCB'17)*, pages 659–665. IEEE, 2017.

[2] E. Alles, Z. Geradts, and C. Veenman. Source camera identification for heavily jpeg compressed low resolution still images. *Journal of Forensic Sciences*, 54(3):628–638, 2009.

[3] I. Amerini, R. Caldelli, P. Crescenzi, A. D. Mastio, and A. Marino. Blind image clustering based on the normalized cuts criterion for camera identification. *Signal Processing: Image Communication*, (29):831–843, 2014.

[4] S. S. Arora, M. Vatsa, R. Singh, and A. Jain. On iris camera interoperability. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 346–352, Sept 2012.

[5] B. b. Liu, H. K. Lee, Y. Hu, and C. H. Choi. On classification of source cameras: A graph based approach. In *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*, pages 1–5, Dec 2010.

[6] N. Bartlow, N. Kalka, B. Cukic, and A. Ross. Identifying sensors from fingerprint images. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 78–84, June 2009.

[7] G. Bloy. Blind camera fingerprinting and image clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(3):532–534, Mar. 2008.

[8] R. Böhme and M. Kirchner. *Media Forensics*, page 231259. Artech House - Boston, 2016.

[9] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti. Fast image clustering of unknown source images. In *IEEE International Workshop on Information Forensics and Security (WIFS) 2010*, pages 1–5, 2010.

[10] K. Choi, E. Lam, and K. Wong. Automatic source camera identification using the intrinsic lens radial distortion. *OPTICS EXPRESS*, 14(24):11551–65, 2006.

[11] N. Damer, Y. Wainakh, V. Boller, S. von den Berken, P. Terhörst, A. Braun, and A. Kuijper. MorGAN: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In *Proceedings of the 9th IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*. IEEE, 2018.

[12] J. Daugman. How iris recognition works. *International Conference on Image Processing*, 1:I–33–I–36, Dec. 2002.

[13] L. Debiasi, N. Damer, A. M. Saladie, C. Rathgeb, U. Scherhag, C. Busch, F. Kirchbuchner, and A. Uhl. On the detection of GAN-based face morphs using established morph detectors. In *Proceedings of the 20th International Conference on Image Analysis and Processing (ICIAP'19)*, Trento, Italy, 2019.

[14] L. Debiasi, C. Kauba, B. Prommegger, and A. Uhl. Near-infrared illumination add-on for mobile hand-vein acquisition. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS'18)*, Los Angeles, California, USA, 2018.

[15] L. Debiasi, C. Kauba, R. Schraml, and A. Uhl. Towards drug counterfeit detection using package paperboard classification. In *Advances in Multimedia Information Processing – Proceedings of the 17th Pacific-Rim Conference on Multimedia (PCM'16)*, Springer LNCS, Xi'an, CHINA, 2016.

[16] L. Debiasi, C. Kauba, and A. Uhl. Identifying iris sensors from iris images. In C. Rathgeb and C. Busch, editors, *Iris and Periocular Biometric Recognition*, chapter 16, pages 359–382. IET, London, UK, 2017.

[17] L. Debiasi, C. Kauba, and A. Uhl. Identifying the origin of iris images based on fusion of local image descriptors and PRNU based techniques. In *Proceedings of the IAPR/IEEE International Joint Conference on Biometrics (IJCB'17)*, Denver, Colorado, USA, 2017.

[18] L. Debiasi, S. Kirchgasser, B. Prommegger, A. Uhl, G. Artur, and K. Marcin. Biometric template protection in the image domain using non-invertible grey-scale transforms. In *Proceedings of the IEEE Workshop on Information Forensics and Security (WIFS'19)*, pages 1–6, Delft, Netherlands, 2019. accepted.

[19] L. Debiasi, E. Leitet, K. Norell, T. Tachos, and A. Uhl. Blind source camera clustering of criminal case data. In *Proceedings of the 7th International Workshop on Biometrics and Forensics (IWBF'19)*, Cancun, Mexico, 2019.

[20] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch. PRNU variance analysis for morphed face image detection. In *Proceedings of the IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS'18)*, Los Angeles, California, USA, 2018.

[21] L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch. PRNU-based detection of morphed face images. In *Proceedings of the 6th International Workshop on Biometrics and Forensics (IWBF'18)*, Sassari, Italy, 2018.

[22] L. Debiasi and A. Uhl. Blind biometric source sensor recognition using advanced PRNU fingerprints. In *Proceedings of the 2015 European Signal Processing Conference (EUSIPCO'15)*, Nice, France, 2015.

[23] L. Debiasi and A. Uhl. Techniques for a forensic analysis of the casia-iris v4 database. In *Proceedings of the 3rd International Workshop on Biometrics and Forensics (IWBF'15)*, Gjøvik, Norway, 2015.

[24] L. Debiasi and A. Uhl. Comparison of PRNU enhancement techniques to generate PRNU fingerprints for biometric source sensor attribution. In *Proceedings of the 4th International Workshop on Biometrics and Forensics (IWBF'16)*, Limassol, Cyprus, 2016.

[25] L. Debiasi and A. Uhl. PRNU enhancement effects on biometric source sensor attribution. *IET Biometrics*, 4(6):256–265, 2017.

[26] Z. Deng, A. Gijsenij, and J. Zhang. Source camera identification using auto-white balance approximation. In *Proceedings of the IEEE International Conference on Computer Vision, ICCV'11*, pages 57–64, Barcelona, Spain, Nov. 2011.

[27] A. E. Dirik, H. T. Sencar, and N. Memon. Digital single lens reflex camera identification from traces of sensor dust. *IEEE Transactions on Information Forensics and Security*, 3(3):539–552, Sept. 2008.

[28] A. E. Dirik, H. T. Sencar, and N. Memon. Flatbed scanner identification based on dust and scratches over scanner platen. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1385–1388, 2009.

[29] N. Erdogmus and S. Marcel. Spoofing face recognition with 3d masks. *IEEE transactions on information forensics and security*, 9(7):1084–1097, 2014.

[30] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *Proceedings of the 2014 International Joint Conference on Biometrics (IJCB)*. IEEE, sep 2014.

[31] M. Ferrara, A. Franco, and D. Maltoni. On the effects of image alterations on face recognition accuracy. In *Face Recognition Across the Imaging Spectrum*, pages 195–222. Springer International Publishing, 2016.

[32] I. O. for Standardization. Iso/iec 30107-1:2016. *Information technology - Biometric presentation attack detection - Part 1: Framework*, 2016.

[33] J. Fridrich. *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, Nov. 2009.

[34] J. Galbally-Herrero, J. Fierrez-Aguilar, J. Rodriguez-Gonzalez, F. Alonso-Fernandez, J. Ortega-Garcia, and M. Tapiador. On the vulnerability of fingerprint verification systems to fake fingerprints attacks. In *Proceedings 40th Annual 2006 International Carnahan Conference on Security Technology*, pages 130–136. IEEE, 2006.

[35] T. Gloe and R. Böhme. The dresden image database for benchmarking digital image forensics. In *SAC 2010: Proc. of the 2010 ACM Symposium on Applied Computing*, pages 1584–1590. ACM, 2010.

[36] M. Goljan, J. Fridrich, and T. Filler. Large scale test of sensor fingerprint camera identification. In *Proceedings of SPIE, Electronic Imaging, Security and Forensics of Multimedia Contents XI*, San Jose, CA, USA, Jan. 2009. SPIE.

[37] M. Goljan, J. Fridrich, and T. Filler. Managing a large database of camera fingerprints. In *Proceedings of SPIE, Media Forensics and Security XII*, San Jose, CA, USA, Jan. 2010. SPIE.

[38] D. Goodin. Fingerprint lock in samsung galaxy 5 easily defeated by whitehat hackers, 2015.

[39] J. Hämmerle-Uhl, K. Raab, and A. Uhl. Robust watermarking in iris recognition: application scenarios and impact on recognition performance. *ACM SIGAPP Applied Computing Review*, 11(3):6–18, 2011.

[40] G. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(3):267–276, Mar 1994.

[41] H. Hofbauer, L. Debiasi, and A. Uhl. Mobile face recognition systems: Exploring presentation attack vulnerability and usability. In *Proceedings of the 12th IAPR/IEEE International Conference on Biometrics (ICB'19)*, Crete, Greece, 2019.

[42] M. Jahanirad, A. W. A. Wahab, and N. B. Anuar. An evolution of image source camera attribution approaches. *Forensic Sci Int*, 262:242275, 2016.

[43] J. R. Janesick, T. Elliott, S. Collins, M. M. Blouke, and J. Freeman. Scientific charge-coupled devices. *Optical Engineering*, 26(8), 1987.

[44] N. Japkowicz and S. Stephen. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5):429–449, February 2002.

[45] N. Kalka, N. Bartlow, B. Cukic, and A. Ross. A preliminary study on identifying sensors from iris images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.

[46] J. Lee. Spoofing iris recognition technology with pictures, 2015.

[47] C.-T. Li. Unsupervised classification of digital images using enhanced sensor pattern noise. In *ISCAS*, pages 3429–3432. IEEE, 2010.

[48] C.-T. Li and X. Lin. A fast source-oriented image clustering method for digital forensics. *EURASIP Journal on Image and Video Processing*, 2017(1):69, 2017.

[49] X. Lin and C.-T. Li. Large-scale image clustering based on camera fingerprints. *IEEE Trans. on Information Forensics and Security*, 12(4):793–808, 2017.

[50] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva. Blind PRNU-based image clustering for source identification. *IEEE Trans. on Information Forensics and Security*, 12(9):2197–2211, 2017.

[51] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva. A deep learning approach for iris sensor model identification. *Pattern Recognition Letters*, pages 1–9, 2017. to appear.

[52] R. C. Pandey, S. K. Singh, and K. K. Shukla. Passive forensics in image and video using noise features: A review. *Digital Investigation*, 19:1–28, 2016.

[53] Q.-T. Phan, G. Boato, and F. G. De Natale. Accurate and scalable image clustering based on sparse representation of camera fingerprint. *IEEE Transactions on Information Forensics and Security*, 2018.

[54] A. Piva. An overview on image forensics. *ISRN Signal Processing*, 2013:Article ID 496701, 2013.

[55] R. Ramachandra and C. Busch. Presentation attack detection methods for face recognition systems: A comprehensive survey. *ACM Computing Surveys (CSUR)*, 50(1):1–37, 2017.

[56] N. Ratha, J. Connell, and R. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, Apr. 2001.

[57] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein. Vision of the unseen: Current trends and challenges in digital image and video forensics. *ACM Computing Surveys*, 43(4):Article no. 26, 2011.

[58] K. Rosenfeld and H. Sencar. A study of the robustness of PRNU-based camera identification. In *Proceedings of SPIE, Media Forensics and Security XI*, volume 7254, pages 72540M – 725408M, San Jose, CA, USA, Jan. 2009. SPIE.

[59] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl. Detection of face morphing attacks based on PRNU analysis. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 1(4):302–317, 2019.

[60] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch. Face recognition systems under morphing attacks: A survey. *IEEE Access*, 7:23012–23026, 2019.

[61] R. Schraml, L. Debiasi, C. Kauba, and A. Uhl. On the feasibility of classification-based product package authentication. In *IEEE Workshop on Information Forensics and Security (WIFS'17)*, Rennes, France, December 2017.

[62] R. Schraml, L. Debiasi, and A. Uhl. Real or fake: Mobile device drug packaging authentication. In *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'18)*, Innsbruck, Austria, 2018.

[63] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the 2015 Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2015.

[64] H. Sencar and N. M. (Eds.). *Digital Image Forensics: There is more to a picture than meets the eye*. Springer Verlag, 2012.

[65] A. F. Sequeira, S. Thavalengal, J. Ferryman, P. Corcoran, and J. S. Cardoso. A realistic evaluation of iris presentation attack detection. In *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, pages 660–664. IEEE, 2016.

[66] D. Shullani, M. Fontani, M. Iuliani, O. Al Shaya, and A. Piva. Vision: a video and image dataset for source identification. *EURASIP Journal on Information Security*, 2017(1):15, 2017.

[67] S. Theodoridis and K. Koutroumbas. *Pattern recognition*. Academic press, 1999.

[68] P. Tome and S. Marcel. On the vulnerability of palm vein recognition to spoofing attacks. In *2015 International Conference on Biometrics (ICB)*, pages 319–325. IEEE, 2015.

[69] P. Tome, M. Vanoni, and S. Marcel. On the vulnerability of finger vein recognition to spoofing. In *2014 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–10. IEEE, 2014.

[70] A. Uhl and Y. Höller. Iris-sensor authentication using camera PRNU fingerprints. In *Proceedings of the 5th IAPR/IEEE International Conference on Biometrics (ICB'12)*, pages 1–8, New Delhi, India, Mar. 2012.

[71] K. Wang, B. Wang, and L. Peng. CVAP: Validation for cluster analyses. *Data Science Journal*, 8:88–93, 2009.

[72] M. Wang and W. Deng. Deep face recognition: A survey. *CoRR - Computing Research Repository*, abs/1804.06655, 2018.

A. Appendix

A.1. Breakdown of Authors' Contribution

This section lists a breakdown of authors' contribution with respect to the papers included in this thesis.

Andreas Uhl is/was the thesis advisor/project leader of Luca Debiasi, Christof Kauba, Bernhard Prommegger, Rudolf Schraml and Heinz Hofbauer. Christoph Busch is the thesis advisor/project leader of Ulrich Scherhag and Christian Rathgeb. Florian Kirchbuchner is the thesis advisor/project leader of Naser Damer and Alexandra Mosegui Saladie. Since the explicit contribution of an advisor/project leader cannot be stated for a single paper, it is omitted in the following breakdown.

	Contribution (in %)											
Publication	Luca Debiasi	Christof Kauba	Bernhard Prommegger	Rudolf Schraml	Heinz Hofbauer	Ulrich Scherhag	Christian Rathgeb	Naser Damer	Alexandra Mosegui Saladie	Elisabet Leitet	Kristin Norell	Theodoros Tachos
L. Debiasi and A. Uhl. Blind biometric source sensor recognition using advanced PRNU fingerprints. In *Proceedings of the 2015 European Signal Processing Conference (EUSIPCO'15)*, Nice, France, 2015	100											
L. Debiasi and A. Uhl. Comparison of PRNU enhancement techniques to generate PRNU fingerprints for biometric source sensor attribution. In *Proceedings of the 4th International Workshop on Biometrics and Forensics (IWBF'16)*, Limassol, Cyprus, 2016	100											
L. Debiasi, C. Kauba, R. Schraml, and A. Uhl. Towards drug counterfeit detection using package paperboard classification. In *Advances in Multimedia Information Processing – Proceedings of the 17th Pacific-Rim Conference on Multimedia (PCM'16)*, Springer LNCS, Xi'an, CHINA, 2016	40	40		20								

	Contribution (in %)											
Publication	Luca Debiasi	Christof Kauba	Bernhard Prommegger	Rudolf Schraml	Heinz Hofbauer	Ulrich Scherhag	Christian Rathgeb	Naser Damer	Alexandra Mosegui Saladie	Elisabet Leitet	Kristin Norell	Theodoros Tachos
L. Debiasi and A. Uhl. PRNU enhancement effects on biometric source sensor attribution. *IET Biometrics*, 4(6):256–265, 2017	100											
L. Debiasi, C. Kauba, and A. Uhl. Identifying iris sensors from iris images. In C. Rathgeb and C. Busch, editors, *Iris and Periocular Biometric Recognition*, chapter 16, pages 359–382. IET, London, UK, 2017	60	40										
L. Debiasi, C. Kauba, and A. Uhl. Identifying the origin of iris images based on fusion of local image descriptors and PRNU based techniques. In *Proceedings of the IAPR/IEEE International Joint Conference on Biometrics (IJCB'17)*, Denver, Colorado, USA, 2017	40	60										
R. Schraml, L. Debiasi, C. Kauba, and A. Uhl. On the feasibility of classification-based product package authentication. In *IEEE Workshop on Information Forensics and Security (WIFS'17)*, Rennes, France, December 2017	10	10		80								
L. Debiasi, U. Scherhag, C. Rathgeb, A. Uhl, and C. Busch. PRNU-based detection of morphed face images. In *Proceedings of the 6th International Workshop on Biometrics and Forensics (IWBF'18)*, Sassari, Italy, 2018	50					20	30					
R. Schraml, L. Debiasi, and A. Uhl. Real or fake: Mobile device drug packaging authentication. In *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'18)*, Innsbruck, Austria, 2018	5			95								
L. Debiasi, C. Kauba, B. Prommegger, and A. Uhl. Near-infrared illumination add-on for mobile hand-vein acquisition. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS'18)*, Los Angeles, California, USA, 2018	40	30	30									

Publication	Contribution (in %)											
	Luca Debiasi	Christof Kauba	Bernhard Prommegger	Rudolf Schraml	Heinz Hofbauer	Ulrich Scherhag	Christian Rathgeb	Naser Damer	Alexandra Mosegui Saladie	Elisabet Leitet	Kristin Norell	Theodoros Tachos
L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch. PRNU variance analysis for morphed face image detection. In *Proceedings of the IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS'18)*, Los Angeles, California, USA, 2018	50					25	25					
L. Debiasi, E. Leitet, K. Norell, T. Tachos, and A. Uhl. Blind source camera clustering of criminal case data. In *Proceedings of the 7th International Workshop on Biometrics and Forensics (IWBF'19)*, Cancun, Mexico, 2019	60									15	10	15
H. Hofbauer, L. Debiasi, and A. Uhl. Mobile face recognition systems: Exploring presentation attack vulnerability and usability. In *Proceedings of the 12th IAPR/IEEE International Conference on Biometrics (ICB'19)*, Crete, Greece, 2019	50				50							
L. Debiasi, N. Damer, A. M. Saladie, C. Rathgeb, U. Scherhag, C. Busch, F. Kirchbuchner, and A. Uhl. On the detection of GAN-based face morphs using established morph detectors. In *Proceedings of the 20th International Conference on Image Analysis and Processing (ICIAP'19)*, Trento, Italy, 2019	45					5	5	30	15			
U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl. Detection of face morphing attacks based on PRNU analysis. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 1(4):302–317, 2019	30					40	30					