Exploring Presentation Attack Vulnerability and Usability of Face Recognition Systems

Heinz Hofbauer[1] • Luca Debiasi[1] • Susanne Kränkl[2] • Andreas Uhl[1]

[1]Multimedia Signal Processing and Security Lab, Paris Lodron University of Salzburg, Austria
{hofbauer, ldebiasi, uhl}@cs.sbg.ac.at
[2]Veridos LLC., Munich, Germany
susanne.kraenkl@veridos.com

November 2, 2020

Abstract

We evaluate commercial face recognition software intended for the use of access control. Most of the systems are to be used with hand held devices (smartphones). The systems under test also contain three stationary systems designed to unlock doors or other secure entrance systems. 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) with different complexity levels and template comparison performance. We contrast the robustness against presentation attacks with the systems usability during regular use, and highlight where current commercial of the shelf systems (COTS) stand in that regard. We examine the results focusing on the tradeoff between acceptance, linked with usability, and security, which usually negatively impacts usability. We also present a first extension of the attacks to systems using the NIR spectrum for imaging. This is mostly limited to stationary systems since they can include dedicated hardware with NIR capabilities. This is their main differentiation to most COTS systems running on smartphones, which do not rely on dedicated hardware. Though exceptions to this already exist, for example in Apple devices. We show that most of the systems are not secure and not user friendly, having huge problems with difficult lighting conditions while only providing the most basic liveness or presentation attack detection capabilities.

Contents

1 Introduction **2**

2 Related Work **2**

3 Presentation Attacks **5**
3.1 Presentation Attack: 2-Dimensional Attacks . . . 5
3.2 Presentation Attack: Masks . . . 5

4 Evaluation of the Systems under Test **6**
4.1 Evaluation of Generic Systems . . . 7
4.1.1 LD only . . . 7
4.1.2 FollowIt . . . 7
4.1.3 OnDevice . . . 9
4.1.4 OnServer . . . 9
4.2 Evaluation of Dedicated Systems . . . 9
4.2.1 iPhone/iPad—FaceID . . . 9
4.2.2 Not NIR . . . 10
4.2.3 NIR Slow . . . 10
4.2.4 Stationary: NIR Fast — Setup and Information . . . 10

5 Extending the Presentation Attacks to the NIR Spectrum **10**

6 Usability **12**
6.1 Time to unlock . . . 12
6.2 Varying illumination . . . 12

7 Lessons Learned and Practical Tips **12**

8 Conclusion **13**

1 Introduction

We evaluated commercial face recognition software intended for the use of access control. These systems range from ones designed for unlocking smartphones, and thus work with commercial of the shelf (COTS) hardware, to stationary access control systems, with dedicated hardware. The main difference is that the imaging capabilities of the smartphones, apart from the Apple devices using FaceID, are limited to the visible spectrum, while most systems with dedicated hardware utilize imaging in the near infrared (NIR) spectrum. Nevertheless, all systems share one common property: they are all closed source and act as a *black-box*, i.e., the applied liveness detection and recognition algorithms are unknown and feedback from most systems is only given in terms of *access granted* or *denied*. However, most systems also give the stage at which the failure occurred, i.e., whether liveness detection or biometric comparison failed. The devices were rather diverse but all worked based on the same principle, images from a user were recorded during biometric enrollment. Biometric verification also recorded an image and was prefaced with a presentation attack/liveness detection step. A successful attempt would unlock the smartphone, or the door in case of stationary systems. The biometric verification and liveness detection were performed either on the device itself, on a server or a combination of both.

The company which provided financial funding also provided the servers, in case they were needed, as well as the software licenses. This has some implications: (1) We can not give the name of the systems due to NDAs; and (2) 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 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. Since we do not have the implementation details of the systems under test we can not comment on the software implementation. We can, and will, however, comment on the way the software presents itself, specifically the cues supposed to aid users (or their absence). We will only comment on those details if the impression is mutual for all testers.

We should note that this is an extension of a prior paper ([1]). We included the systems from the prior paper as well. However, we will not draw direct comparisons. Mostly because the implementation of the systems, and the tasks required to unlock the device changed so dramatically that a direct comparison would be pointless.

Limited Tests: Due to time constraints, as mentioned above, we could only afford a limited number of tests. Specifically, most test were only performed by a single, but not necessarily the same, user. Overall, the group involved in the tests consisted of ten users ranging in age from approximately18 to 40 and containing members of both sexes. The security feature can be considered broken if we can show one successful attack, in essence a proof by falsification. Even with a single user we found ample counter-evidence regarding the security of most systems and could formulate a method of attack.

For the usability the limit to a single test user is more of a problem, i.e., a single sample from the population rather than a proper statistical cross section. We performed baseline tests with multiple users for one system, which resulted in a negligible variance for the unlock attempts. However, given that mostly the usability turned out to be bad it shows there still are problems to be solved.

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. We recorded the time for all genuine unlock attempts, usually 10, and divided this by the number of successful unlocks to get an average time to unlock.

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. The presentation attack artifacts, and how they were created, is described in Section 3. Then follows a section of tests, split per device, in Section 4. The grouping of devices is difficult, since most have little to nothing in common. Further, the number of tests makes condensing the information, in a still legible way, difficult. We have split the evaluation into two sections, one dealing with the generic approaches, mostly smartphone based systems, in Section 4.1, and one for systems with dedicated hard- and software, in Section 4.2. The test section follows a common setup to make all systems easier to compare. The shift to NIR imaging, mostly by devices with dedicated hardware, made it necessary to show that the same attacks can be mounted in the NIR spectrum. This is given in Section 5. Section 6 discusses the usability of the systems under various illumination conditions. Finally, Sections 7 and 8 will summarize our findings and conclude the paper.

2 Related Work

Smartphones are ubiquitous and so is the widespread adoption of biometric characteristics to unlock the device by verifying the identify of the user. Primarily fingerprint based systems are currently the norm and dedicated sensors are built into practically every smartphone. In recent years, a trend towards face detection can be observed. These are mostly software based systems because dedicated hardware for face detection is not yet integrated widely into smartphones, with some exceptions like the recent iPhone. This trend has renewed the interest in attacks, and the prevention thereof, against such biometric systems. An overview of the relevant, to the paper, are given in the following and a very brief overview is given in Table 1.

A specific attack is the presentation, also known as direct or spoofing, attack. It can be separated into two categories [23]: (1) active impostor presentation attacks, where the attacker tries to claim a foreign identity; and (2) concealer presentation attacks,

Table 1: Overview of related work grouped by type, i.e., presentation attack detection (PAD), presentation attack instrument (PAI). If applicable the specifics are briefly given, these are attack instrument for PAI and the features used for PAD. Only works from literature relevant to the systems under test are presented, for a more all-encompassing overview we also list surveys which the so-inclined reader can peruse.

Type	Specifics	Reference
PAD	Behavior, blink detection	Gang *et al.* [2] (2007), Chrzan [3] (2014)
PAD	Behavior, challenge response	Kollreider *et al.* [4] (2008), Ali *et al.* [5] (2013), Smith *et al.* [6] (2015)
PAD	Image, texture based features	Kose and Dugelay [7] (2012), Chingovska *et al.* [8] (2012), Yang *et al.* [9] (2013), Boulkenafet *et al.* [10] (2016)
PAD	Image, distortion modelling	Wen *et al.* [11] (2015)
PAD	Image, multispectral fusion, generalization evaluation	Chingovska *et al.* [12] (2016)
PAD	Video based dynamic texture	Tiago de Freitas Pereira *et al.* [13] (2012)
PAD	Video, movement based features	Marsico *et al.* [14] (2012), Anjos *et al.* [15] (2014), Pinto *et al.* [16] (2015)
PAI	Images, from social networks (facebook)	Li *et al.* [17] (2018)
PAI	Makeup (no PAD was used)	Chen *et al.* [18] (2017)
PAI	Masks, comparison of VIS and NIR	Agarwal *et al.* [19] (2017), Ramachandra *et al.* [20] (2019)
PAI	Masks, Print, Image, Video (preliminary work to this)	Hofbauer *et al.* [1] (2019)
Surveys		Chingovska *et al.* [12] (2016), Jia *et al.* [21] (2020), Jia *et al.* [22] (2020)

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 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 biometric characteristics, parts of dead bodies, modified faces, forced presentation by unconscious persons and so on.

To prevent presentation attacks, a presentation attack detection (PAD) system, also referred to as liveness detection, is employed. The primary focus of research are artificial presentation attack instruments but, as is evident in the term liveness detection, overlaps with parts of the human biometric characteristic PAI categorization. There are different kinds of (face detection) PAD methods, 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 ([7–10]), dynamic texture based (video) ([13]) or movement based ([14–16]). For more details, the reader is referred to the respective papers.

This is of course not a comprehensive list, and we omitted state of the art techniques which could not have been employed in the devices under test. Most of the commercial of the shelf systems do only have a single visible light camera, so multispectral PAD methods can not be employed. Similarly, DNN based methods can not be employed since the devices lack the required computational capacity. For a more thorough overview, the reader is directed to [12] for a survey-like overview of (mostly) 2D PADs and [21] and for a more recent state of the art overview. In [22] the authors give a concise overview of PADs for 3D mask presentation attacks.

The target application of our tests was to unlock the device with the presented biometric characteristic (face). For smartphones the operation mode, in terms of biometry, is always verification since the identity is implied (the owner of the cell phone). Presentation attacks also try to unlock the device and are consequently also done in verification mode. For the stationary systems the target application is a biometric access control, so identification is required. However, some systems require the user to identify themselves, but some allow for verification by claiming an identity using a unique code, which can optionally be stored on a NFC enabled device, e.g., an ID card. 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.

Please note that there are other similar studies in which commercial of the shelf (COTS) face based biometric systems have been evaluated. None of them took into account usability, user experience and consequently user acceptance. Usually, papers in literature use software on databases and therefor preclude a more interactive creation of the presentation attack instrument which, instead they simply use the recording of the presentation attack from a database. Most of them do not take into account NIR imaging and/or the specific impact of lighting on the PAD systems.

There is a preliminary work [1] of which the current paper is an extension. The paper dealt with an exploratory look into two face biometric systems. The outlines established there were used as a basis for this work. Having guidelines what to test and who to test allowed us to increase the number of systems under test, from two systems on one device to four systems on four different devices and five dedicated systems. The current work

also includes the systems under test in [1], although they have been updated to such an extent that they can be considered new systems.

In [20] the authors tested expensive silicone masks on two commercial face recognition systems, and also on learning based (SVM) PADs from literature. The main conclusion was that the face recognition places the masks closer to genuine than true imposter attempts, not so near that they are not cleanly separable. Some of the PADs could learn to identify the masks if trained with images from the same device (three different smart phones were used). However, transfer learning between the devices did poorly. Usability was not taken into account, neither were different environmental conditions (like lighting).

Another paper dealing with masks is [19], in which the near infrared (NIR), visible light (VIS) and thermal spectrum are compared in relation to presentation attack detection of masks. The goal is to detect obfuscation so no attack on face recognition systems is considered. The PAD is based on different features from literature which are used in SVM-based machine learning. No transfer learning between spectra or devices is examined, and neither is fusion between the different spectra. They key results from the paper are that PAD detection of masks does not work very well in the NIR spectrum (best EER of 42%), slightly better in the VIS spectrum (best EER of 29.2%), and only decently well in the Thermal spectrum (best EER of 10.8%). These results are for video based detection, still image based detection is worse in every case.

In [18] makeup instead of masks was used. Only one commercial face recognition system was used, but also one from literature, a VGG-Very-Deep-16 architecture convolutional neural network [24]. They showed that makeup can be potentially used to spoof face recognition systems. No PAD systems were considered in this work.

In [11] the authors introduce a new PAD method based on image distortion models, specifically reflection, blurriness and color diversity. A SVM was used to learn the presence of presentation instruments. They tested on three databases, containing images from a uniformly lit interior environment, but with very different sensors. The tests showed that their distortion based model for the most part is better than the texture based methods. They evaluated intra and inter database training/classification showing that inter database, and therefor sensor, classification is harder than inter database classification. More specifically, they looked at the cameras used to record the images in the databases and found that imaging devices of similar quality lead to a better transfer of the learned models than different imaging devices.

The threat of using face images leaked via social networks by users themselves is evaluated in [17] and found to be quite severe. They also differentiated between low and high security modes, with high security modes entailing a higher rejection rate of genuine attempts while reducing the attack success rate. They found specifically that high quality unaltered images pose the highest threat while blurry, low resolution, edited or images of the user wearing makeup can degrade the attack value of the image. They also compared the tolerances in head pose and environmental lighting conditions, showing that low security settings generally tend to be a lot more tolerant. The lighting evaluation is only using limited natural lighting environment. Here our experiments, especially in the controlled environment, are more thorough. However, the basic trend is also confirmed in [17], namely that the methods work best in a well lit indoors environment and get very much worse in every other lighting environment. They evaluated the basic liveness detection and used image manipulation to create interactive videos from the images to simulate liveness, this is somewhat similar to what we do with paper masks but digitally. The results show that false rejection rate is increase but false acceptance rate is decreased, but not to zero.

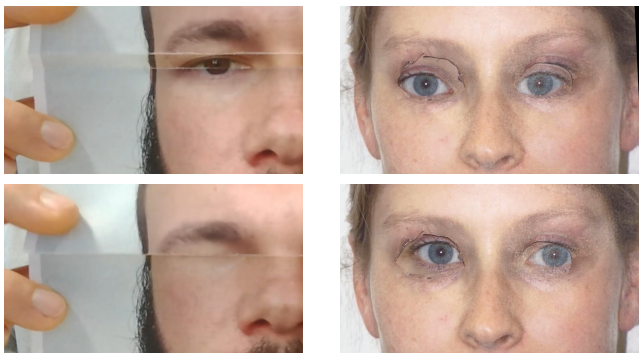
A comprehensive overview and comparison of current state of the art presentation attack detection systems is performed in [21]. Their main complaint about surveys was that the reported performances are from the original papers and, using different protocols and databases, are not comparable. They therefor took 30 PAD systems from literature and ran the same tests on three publicly available databases to generate a proper comparison and evaluation. The showed that there is ample room for improvement for PAD systems. Overall the results are very mixed and no single method was the clear 'best' option over all databases, however, as a class deep learning based methods clearly performed very well. Further, they showed that the more changes from one image to the next, in terms of lighting, resolution, background and so on, the harder it was to correctly identify presentation attacks. They also evaluated transfer learning, i.e., training the algorithm on one database and evaluating on another, with the same result as other works, while possible the results on different databases are strongly degraded. No systematic evaluation of the effect of lighting conditions was taken into account and all tests were performed on visible light only.

In [12] the authors look at mask based spoofing and multispectral fusion. Main conclusions are that multispectral fusion can somewhat alleviate the threat of attack but not prevent it. Further, they found that machine learning for mask based presentation attack detection poorly generalizes, such that even masks with similar manufacturing methods can at times totally bypass such a PAD system. The paper also gives a very good survey-like overview of spoofing counter measure (presentation attack detection).

A newer survey [22] presents a unified look at 3D mask PAD systems in a similar way to [21]. Specifically the paper presents a good overview of 3D (and 2D) PAD systems and 3D Mask databases. The authors then choose 10 PAD systems (those for which source code was available from the original authors) and subject them all to the same tests over two databases. Main results are that some classical 2D texture based PAD systems, like multiscale LBP, are very able to detect 3D Masks. For a simple database containing only one type of mask (from ThatsMyFace) deep-learning based systems were also able to detect the masks. For a more complex database, containing masks from a different manufacturer, the deep learning based methods performed significantly worse than the classical methods. But both showed a severe degradation in performance showcasing the poor generalization performance (basically similar result to [12] and [21]). Only visible light imaging was taken into account.

Table 2: 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	video of the face	3D face reconstruction



(a) Folding a pre-creased sheet of paper to simulate blinking.

(b) Two layered prints to allow simulation of eye movement.

Figure 1: Various interactive attack vectors using paper prints.

3 Presentation Attacks

We differentiate between replay attacks, which use an external imaging source to record an image and replay that image to the sensor. This can be from a screen or print or by creating a paper mask. The second attack type is more complex and generates a wearable, more realistic looking, presentation attack artifact in the form of a mask. This second form of attack is especially important in conditions where the user might be observed, holding sheets of printed paper in front of ones face is quite obvious, wearing a realistic lifelike mask less so. We will also give information about the complexity of mounting the described attack. To do this we will use the threat level model laid out in [25] and summarized in Table 2.

3.1 Presentation Attack: 2-Dimensional Attacks

These types of attacks use flat presentations, as in there is not depth information, for the attack. It includes the presentation of printouts or displays (photo attack) or the presentation of videos on a 2-dimensional screen (replay attacks). 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. We used the following types of presentation attack artefacts for replay attacks:

Print For this presentation attack (PA) a print of the image on paper was used. The print is in color and of a sufficiently high quality.

Screen This PA displays the captured image on a screen, e.g., tablet, laptop, or similar.

Video An extension of Screen, in that it replays a recorded video of the subject. This PA was only done via a laptop.

Three main types of interaction were present across the systems: blinking, eye tracking and head tracking. It should be noted that Print especially was used for interactive modes in various ways. While Video seems the more obvious choice, the interaction with the device was not static but slightly randomized. So an attack where the user could interact with the device was simpler to mount with a sheet of paper which could be manipulated in real time as opposed to a video. For head tracking, all attack types could be used as a simple sideways shift of the PAI was usually sufficient to fool the system. Interactivity was only the direction of the shift which could be accommodated with all PAIs. This was not the case for blinking and eye tracking which required an interaction of a part of the PAI (eyes) rather than the whole (head).

Interaction with blinking-based liveness detection was achieved by folding a paper artefact across the eyes. This way the eyes can be hidden, when the paper is folded, or shown when unfolded, allowing for a simulated blinking. Figure 1a shows the crease along which the paper is folded as well as the folded 'blink' simulation. A different solution was used when eye-tracking was involved in the interactive unlocking process. Here, two prints of the same face were used, layered on top of one another. The top sheet had the eyes cut out, this way the bottom sheet (including the eyes) can be moved independently to the rest of the 'face' simulating eye movement similar to looking into a certain direction, Fig. 1b illustrates the process.

Complexity of Attacks

The complexity of these attacks is extremely basic. The images we used where taken with a smartphone. The following screen-based presentation attacks also utilized a smartphone. The Print attacks have a slightly higher complexity in that they require an additional device, a color printer. Given the ubiquitousness of all involved devices we have to conclude that these attacks can be mounted by everyone, no special knowledge or equipment is required. Clearly a threat of 'Level A'.

A video is slightly more difficult to obtain, not because special knowledge or hardware is required, but simply because the subject has to be recorded over a longer period of time. Alternatively, a specialized (small) device could be placed somewhere where the recording is simple to obtain. This slight increase in difficulty pushes the threat to 'Level B', as at least some planning is usually required.

3.2 Presentation Attack: Masks

For these presentation attacks, we used a mask or mask-like presentation of the stolen biometric characteristic, 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 three attack types, shown in Fig. 2:

Mask—Paper Similar to Print this is a printed version of the image, with cutout eyes and nose so it can be worn.

Mask—Latex A handcrafted 3D latex mask by CREA FX[1]

Mask—Resin A 3D-printed hard resin composite mask by ThatsMyFace[2].

It should be noted that the masks all had a problem, which is the illumination of the eyes. The masks add height to the face, which in turn recesses the eyes and darkens them, often leading to problems during the presentation attack, especially for modes where eye-tracking was involved. This could be solved by providing better illumination. This solution is simple for the unlocking of mobile devices but can be problematic with stationary systems.

Complexity of Attacks

The creation of the paper mask is simple, and basically the same as a paper print. However, the illumination problem makes the application of the attack more of a problem and can push the threat for Mask—Paper from 'Level A' to 'Level B' in case of stationary systems. The Mask—Latex and Mask—Resin artefacts are much more difficult to mount. The masks cost a substantial amount of money and the production takes some time. In addition the requirement for even illumination of the face might make attacks more obvious which in turn requires more planning. This puts the Mask—Latex and Mask—Resin presentation attacks squarely into threat 'Level C'.

4 Evaluation of the Systems under Test

Before going into the specifics of the test we briefly present an overview over the tests, systems under test and the devices we tested the systems on. The section on a specific system will have more information about the system and also discussion of results or remarks specific to that system.

For usability evaluation we tested how long it takes for a genuine user enrolled on the device to unlock the device. We will record the **time to unlock (TTU)**, in seconds, as measure for usability, in the sense that the user satisfaction (or inversely frustration) is directly tied to the amount spent to unlock the device. It should take only briefly (<5s) to unlock the device, everything above that is intrusive. The TTU is the time it takes for all unlock attempts divided by the successful unlocks, i.e., it takes failures into account. The usability tests were performed with natural and artificial light. The natural light test was done as a baseline, the light conditions varied of course which makes for poor comparability. Therefore, we also used a controlled light setup as a control experiment. For the controlled light setup we utilized 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 80g/m^2 paper) light to simulate a clear or cloudy day respectively.

[1] https://www.creafx.com/en/
[2] http://thatsmyface.com/custom-wearable-masks/

Table 3: Overview over the experiments. For each systems, and configuration (mode) where applicable, the number of devices (#dev) and the number of presentation artefact instruments (PAI) per device are given.

(a) Generic Systems

System	mode	#dev	PAI
LD ONLY		3	6
FOLLOWIT		2	4
ONDEVICE	Active	2	4
ONDEVICE	Passive	2	5
ONSERVER	lenient	3	6
ONSERVER	strict	3	6

(b) Dedicated Systems

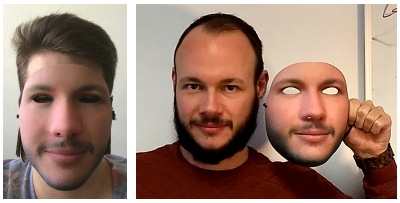
System	#dev	PAI
FACEID	2	6
NOT NIR	1	6
NIR SLOW	1	6
NIR FAST	1	6

For the security evaluation we look at two different results. One is the success of overcoming the liveness detection and the other is the actual comparison of the presented biometric probe to the stored biometric reference. In the following they will be given in percent and denoted **liveness detection attack (LDA)** and **biometric verification attack (BVA)** rate. Some systems we used only supported liveness detection, while most also supported biometric verification. The majority of the systems under test use a staged system of liveness detection before biometric verification, meaning that the BVA rate can not be higher than the LDA rate. If a system deviates from this general setup it will be explicitly stated in the relevant system description.

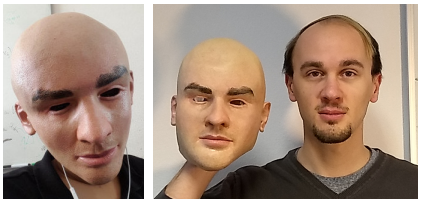
We differentiate between dedicated systems (Section 4.2), with custom hard- and software, and generic systems (smartphones), which contain different hardware components and a software which has accommodate this variation in hardware (Section 4.1). Further, we used an iPhone which has custom built hard- and software, and is therefore better fitting into the dedicated system despite being a smartphone.

The tests in the following will be presented in a unified fashion, that is for the same test setup the same reporting table will be shown, even if there is some information missing. This should facilitate the comparison of different systems. The generic systems are all comparable in this way, the stationery are different from each other and the generic systems. An overview over the systems under test, differentiated by their configuration if applicable, the number of devices it was tested on and the number of presentation attack artefacts used per device are given in Table 3. Due to NDAs we can not name the systems.

Nota Bene: In the following sections we will present the results in the form of success rate of the liveness detection attack (LDA) and the biometric verification attack rate (BVA) in percent. This relates to the reporting as specified in ISO/IEC 30107-3 [26] as follows: In case the presented biometric characteristic was genuine, the bona fide presentation classification error rate BPCER := $1 - {}^{\text{LDA}}/_{100}$, in case of a presentation attacks the attack presentation classification error rate APCER := $1 - {}^{\text{LDA}}/_{100}$. Like-



(a) Resin mask, the sources of the biometric characteristics (right) hold the replica and imposter wearing it (left).



(b) Latex mask, the sources of the biometric characteristics (right) hold the replica and imposter wearing it (left).

(c) Paper masks, prints with eyes removed. The nose can also be removed to make the mask easier to wear (right).

Figure 2: Wearable presentation attack artefacts (also known as masks).

Table 4: Overview over camera and CPU in the smartphones used for the generic systems tests.

Phone	Main Camera	CPU
LG K8	8 MP, f/2.4, AF[1]	Quad-core: 4x1.3 GHz
SXC 4	13 MP, f/1.9, AF[1]	Quad-core: 4x1.4 GHz
SG S8	12 MP, f/1.7, 26mm (wide), PDAF[2], OIS[3]	Octa-core: 4x2.3 GHz + 4x1.7 GHz
O+ 6	16 MP, f/1.7, 25mm (wide), PDAF[2], OIS[3]	Octa-core: 4x2.8 GHz + 4x1.7 GHz

[1] Auto Focus
[2] Phase Detection Auto Focus
[3] Optical Image Stabilization

wise, the false non match rate FNMR := $1 - {}^{\text{BVA}}/_{100}$ for genuine presentation and the impostor attack presentation match rate IAPMR := ${}^{\text{BVA}}/_{100}$.

4.1 Evaluation of Generic Systems

For this evaluation we used a mix of older and newer smartphones. This changes calculation speed and camera quality, giving a good cross section over conditions a generic software has to deal with. While not all systems could be made to work with each smartphone, we still included each smartphone in each result table. All the systems have a similar design goal and the same restrictions and should therefor be considered directly comparable. However, the camera resolution and quality, due to age of hardware, is widely different. So are the processor capabilities which could lead to large differences in the TTU. The smartphones used are: LG K8 (LG K8), Samsung XCover 4 (SXC 4), Samsung Galaxy S8 (SG S8), and OnePlus 6 (O+ 6). The relevant hardware capabilities are summarized in Table 4. Liveness detection and biometric verification were mostly performed on the device, exception will be noted.

Most systems gave hints to the user where to put the face. This usually took the form of an oval, with the instruction to place the camera such that the face was inside the oval and of the same size. This is a simple way of bringing position and size close to a known value. Some systems instead showed a rectangle where the face was detected, usually in conjunction with a score. This also allows a user to optimize the placement of the face. However, the user has to learn the correct position by trial and error, i.e., moving the head and or camera around and monitor the score. Clearly, the first method is preferrable from a usability point of view.

4.1.1 LD only

This system does not perform biometric verification, only liveness detection. The user is asked to show his face frontally and rotate the head left and right randomly. The final result is a liveness score.

Results for the experiments are given in Table 5a. The first thing to note is that the unlocking procedure is slow due to the interactive nature. This combined with the large success rate of attacks makes this system unusable. Regarding the attacks, it is of little surprise that naturally 3-dimensional presentation attack artifacts can easily deal with this liveness detection method. The inability of the video to do the same is clear, since the tilting of a head can be recorded but this recording will almost certainly be in a different order than specified by the system. What is more surprising is that simply tilting the paper printout was sufficient to fool the system in 90% of the cases. What is even more surprising is that the same presentation attack is not possible with a screen presentation of the image. Unfortunately we could not explain why this is so.

When it comes to lighting condition sensitivity this method overall performs quite well. It has problems in some natural light conditions, especially with front light. But it is one of the few systems that can handle back light well. The usual unlock scenario for smartphones is to look down into the phone. This positions in most cases the illumination, be it ceiling lights or the sun, at the back of the user. So handling this situation well is important.

4.1.2 FollowIt

This is an interesting system in that it performs the liveness detection and biometric verification in an unusual order. The system first performs biometric verification, with the usual visual cues, and then liveness detection followed by another biometric verification step. The liveness detection shows a simple visual cue, which randomly moves across the screen, that the user is supposed to follow with his/her eyes.

Results of the experiments are given in Table 5b. Basically, every type of attack that can simulate eye-movement in a non-static fashion can easily circumvent the liveness detection. This means, screen and video based methods won't work, the prerecorded eye-movement does not match the interactive cues. A regular paper print also will not work of course, but a simple adaptation, as

Table 5: Evaluation results for Generic Systems. Liveness detection attack (LDA), biometric verification attack rate (BVA) and time to unlock (TTU) given for the systems under test.

(a) LD ONLY – Evaluation Results

	O+ 6		SG S8		SXC 4		LG K8	
	LDA	BVA	LDA	BVA	LDA	BVA	LDA	BVA
Nat. Light	100		90		80			
TTU	16s		19.7s		39.3s			
Front	60		80		0			
Side	70		100		90			
Back	100		100		100			
TTU	22.5s		16.1s		21.5s			
Print	90		90		100			
Screen	0		0		20			
Video	0		10		0			
Mask - Print	100		100		100			
Mask - Resin	100		100		100			
Mask - Latex	100		100		100			

(b) FOLLOWIT – Evaluation Results

	O+ 6		SG S8		SXC 4		LG K8	
	LDA	BVA	LDA	BVA	LDA	BVA	LDA	BVA
Nat. Light			100	100			100	100
TTU			13s				13.3s	
Front			90	90			90	90
Side			60	60			80	80
Back			0	0			0	0
TTU			15s				18.2s	
Print			100	100			100	100
Screen			0	0			0	0
Video								
Mask - Print			100	100			100	100
Mask - Resin			100	100			100	100
Mask - Latex								

(c) ONDEVICE: Active – Evaluation Results

	O+ 6		SG S8		SXC 4		LG K8	
	LDA	BVA	LDA	BVA	LDA	BVA	LDA	BVA
Nat. Light			100	100			100	100
TTU			15.2s				13.8s	
Front			100	100			40	40
Side			0	0			0	0
Back			0	0			0	0
TTU			23.9s				38.3s	
Print			70	70			60	60
Screen								
Video			40	40			90	90
Mask - Print			100	100			70	70
Mask - Resin			0	0			0	0
Mask - Latex								

(d) ONDEVICE: Passive – Evaluation Results

	O+ 6		SG S8		SXC 4		LG K8	
	LDA	BVA	LDA	BVA	LDA	BVA	LDA	BVA
Nat. Light			100	100			100	100
TTU			18.8s				17.5s	
Front			100	100			0	0
Side			0	0			0	0
Back			0	0			0	0
TTU			28s				∞	
Print			80	80			80	80
Screen			90	90			100	100
Video			100	100			100	100
Mask - Print			100	100			80	80
Mask - Resin			0	0			0	0
Mask - Latex								

(e) ONSERVER: lenient – Evaluation Results

	O+ 6		SG S8		SXC 4		LG K8	
	LDA	BVA	LDA	BVA	LDA	BVA	LDA	BVA
Nat. Light	100	100	100	100	100	100	0	0
TTU	12.3s		13.0s		12.5s		∞	
Front	60	60	40	40	10	10		
Side	50	50	60	60	60	60		
Back	30	30	0	0	10	10		
TTU	35s		22.9s		118s			
Print*	100	100	100	100	100	100		
Screen	0	0	0	0	0	0		
Video	0	0	0	0	0	0		
Mask - Print	0	0	0	0	0	0		
Mask - Resin	0	0	0	0	0	0		
Mask - Latex	0	0	0	0	0	0		

(f) ONSERVER: strict – Evaluation Results

	O+ 6		SG S8		SXC 4		LG K8	
	LDA	BVA	LDA	BVA	LDA	BVA	LDA	BVA
Nat. Light	90	90	90	90	50	50	20	0
TTU	16.1s		18.1s		47.2s		∞	
Front	0	0	30	30	10	10		
Side	20	20	40	40	40	40		
Back	0	0	0	0	0	0		
TTU	50.3s		43.3s		116s			
Print*	100	100	100	100	100	100		
Screen	0	0	0	0	0	0		
Video	0	0	0	0	0	0		
Mask - Print	0	0	0	0	0	0		
Mask - Resin	0	0	0	0	0	0		
Mask - Latex	0	0	0	0	0	0		

presented in Fig. 1b, is sufficient to pass the liveness detection, as the apparent gaze can be adapted to the given cues interactively.

The system is susceptible to strong back light, resulting in a 100% failure rate. Most likely this is because strong back light will cast strong shadows into the eye sockets making the eye tracking impossible. For biometric verification, the strong back light casts the rest of the face into shadow and or can produce glare in the camera lens. Most systems have problem with back light, so this is not surprising. The device overall functions well in natural light conditions otherwise.

4.1.3 OnDevice

This system has two different modes for liveness detection, an active and a passive one. We decided to evaluate both modes separately. Experimental results are given in Table 5c and Table 5d for the active and passive modes respectively.

When it comes to light, both methods and biometric verification could not work with strong side- or back-light. Nonetheless, it seems to work well in natural lighting.

OnDevice—Active: This liveness detection method requires the user to look at the screen and then rotate the head left and right following the system's random instructions. At the end of each instruction, the user is asked to blink. This process is repeated mulitple times for a single liveness detection check.

A screen attack could not work since blinking was not possible. A regular print attack also suffered from the same problem, but given the physicality of the print we could simulate blinking by folding the paper at the appropriate time, see Fig. 1a. The sideways movement was simulated simply by shifting the presentation artifact left or right. For the video attack the head movement was performed in the same way. The video sequence contained blinking, but the timing of when the blinking should occur, due to the cue, and when it actually occurred, due to the timing in the video, was hard to synchronize. This overall lead to a less successful number of attacks than the paper artifact. An additional problem with the video attack was the illumination of the smartphone, which reflected in the device used for presentation. This could be solved by reducing the brightness of the smartphone. The masks had the problem that the eyes had to be well illuminated for the blinking to be registered. Even with proper frontal lighting the resin and latex masks never could pass liveness detection. It is not clear why, since the paper mask could successfully be used to attack the system. For the paper mask the nose had to be cut out, likely so the mask could be worn closer to the face, and consequently the eyes could be better illuminated. The lack of 3-dimensionality, without the cut out nose, should not be an issue given that a paper print could also bypass the system.

OnDevice—Passive: This mode simply required the user to keep still in front of the device. We found that a slight movement helped in circumventing the liveness detection. It is not quite clear what this mode does in the background but it is very bad at detecting presentation attacks. Interestingly even in this mode the latex and resin masks could not penetrate the system, while all other attacks had a high success rate.

4.1.4 OnServer

This system had two modes, one more strict (*strict*) than the other (*lenient*). It is unclear what the difference is internally. Both modes require the user to present the face centrally. No user interactions are required, the user is only informed about the liveness detection and biometric verification results. The liveness detection preceded the biometric verification and both were done on a remote server. Disregarding the results, the reliance on a solely server based system has implications when traveling in a dead zone, the server has a less than 100% reliability, or when the server is breached. The results are given in Tables 5f and 5e for the strict and lenient modes respectively.

While the lenient mode worked well in natural light, we already experienced some failed attempts for the strict mode. The number of failed attempts significantly increased for both modes with varying illumination. The successful attempts using the lenient mode decrease to approximately 50% for light from the front and side, while light from the back shows even worse results. For the strict mode, all types of illumination changes lead to most verification attempts failing. Similar to others, this system requires the facial portion of the image to exhibit a minimum amount of contrast. Too much or too less contrast triggers an error indicating that the quality is insufficient and leads to a failed verification attempt.

Considering the various presentation attacks, both modes look quite robust at first glance. The only successful attack was made with a printed face image, which worked consistently though and was easy to reproduce. However, it has to be noted that the print attack did not work for any tested subject. The print attack was only successful for subjects with lighter skin type and bright hair, while darker types were not attacked successfully. We were not able to perform further investigations in regard to this special behavior due to time constraints.

4.2 Evaluation of Dedicated Systems

Dedicated systems are those where dedicated hardware and software is available. We further differentiate the systems by type/size into stationary systems and the iPhone/iPad based FaceID. The stationary systems are comparable, in that they have a relatively large enclosure in which an ample amount of dedicated hardware can be installed. The iPhone is much more limited in this regard and more comparable with the other smartphones, except that it does indeed contain dedicated hardware, making it a special case. Liveness detection and biometric verification were performed on device for all dedicated systems.

4.2.1 iPhone/iPad—FaceID

The device uses a special hardware, a NIR dot projector, to get a 3D model of the face. Internally a machine learning based model of the face is generated and updated over time to accommodate changes in appearance.

The experimental results are summarized in Table 6a. This systems works as intended and unlocks the system fast and reliably. In the limited time we had to evaluate the system we could not attack it successfully.

Table 6: Evaluation results for the dedicated systems. Since the iPhone/iPad have different spatial restriction when it comes to dedicated hardware they have been split from the larger (stationary) systems.

(a) FaceID – Evaluation Results

	iPhone X	iPad Pro
	LDA & BVA	LDA & BVA
Nat. Light	100	100
TTU	2.4s	2.2s
Front	100	100
Side	100	100
Back	100	100
TTU	2.5s	2.2s
Print	0	0
Screen	0	0
Video	0	0
Mask - Print	0	0
Mask - Resin	0	0
Mask - Latex	0	0

(b) LD only – Evaluation Results

	Not NIR		NIR Slow		NIR Fast	
	LDA	BVA	LDA	BVA	LDA	BVA
Nat. Light	100	100	100	100	100	100
TTU	13.7s		9.3s		2.1s	
Front	0	100	100	100	100	100
Side	0	100	100	100	100	100
Back	60	100	100	100	100	100
TTU	32.2s		8.9s		2.3s	
Print	100	100	0	0	0	0
Screen	100	100	0	0	0	0
Video	100	100	0	0	0	0
Mask - Print	100	100	0	0	0	0
Mask - Resin	100	100	0	0	0	0
Mask - Latex	100	0	0	0	0	0

4.2.2 Not NIR

This stationary system uses multiple cameras, in the visible light (VIS) and NIR spectrum, and corresponding illumination in the form of LED arrays. The device dynamically adjusts illumination and camera orientation. From what we could gather the liveness detection seems to be continuous, and biometric verification is done once the liveness detection has been passed.

Results are summarized in Table 6b under the column Not NIR. The device does not seem to use the NIR imaging capabilities for biometric verification or liveness detection. It could very easily be attacked and stronger light could easily confuse the sensor, another indicator that the NIR camera is underutilized.

4.2.3 NIR Slow

This system uses two cameras, one NIR and one visible spectrum, respectively. The device has six 850nm NIR LED lights in two groups, see Fig. 4a. The face is detected anywhere in the field of view of the camera, an LCD screen shows the camera perspective.

Experimental results are shown in Table 6b under the NIR Slow column. The device uses the NIR spectrum for liveness detection and biometric verification and consequently is unaffected by the strong lights. Due to the sensing in the NIR spectrum the regular (VIS) attacks fail liveness detection. However, preliminary NIR attack tests show that the liveness detection is not very strong, see Section 5 for more information about these attacks.

4.2.4 Stationary: NIR Fast — Setup and Information

Like the previous systems, two cameras are used, one NIR and one visible spectrum camera. Two LED arrays provide illumination in the 850nm frequency range, see Fig. 3b. The face can be placed anywhere in the field of view of the camera with feedback shown on an LCD screen.

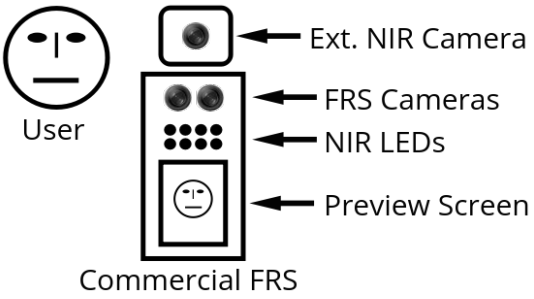
Evaluation results are shown in Table 6b in the NIR Fast column. Like before, the NIR imaging means that the regular attacks mounted in the visible spectrum do not work, and that the device is unaffected by surrounding light. Like with NIR Slow, the attacks in the NIR spectrum work well, showing that the liveness detection is not very sophisticated, again see Section 5 for more information about this. A positive of this system is the speed of liveness detection and biometric verification.

5 Extending the Presentation Attacks to the NIR Spectrum

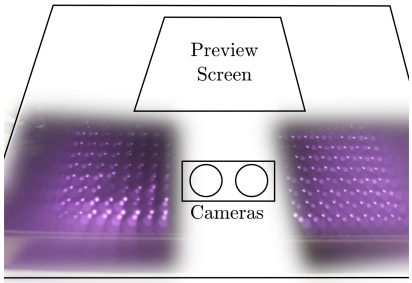
The presentation attacks so far have been mounted in the visible spectrum. The primary countermeasure, at least for commercial systems, seems to be a shift to the NIR spectrum if possible, i.e., in dedicated hardware. The question is then, is this simply a shift in spectrum or is there also a better liveness detection at work.

If the algorithms for liveness and presentation attack detection are as flawed as in the visible spectrum, the shift to the NIR spectrum will not stop a serious attack. Towards this end we simply shift the acquisition of imagery, used in the presentation attacks, into the NIR spectrum as well. A further question is the reproduction of the acquired image in such a way that it looks the same in the NIR spectrum. We know however from literature that at least the black toner used in laser printers is also black in the NIR spectrum, see [12, 27, 28]. We used a NIR sensitive industrial camera which is commonly used for image acquisition in other biometric systems. This camera has a lower resolution and closer focal point, so a proper attack with an image acquired at a distance could not be performed. However, as a proof of concept this should be sufficient. We mounted the external camera close to the biometric sensor, Fig. 3a and acquired an image using the illumination used by the biometric sensor itself, Fig. 3b.

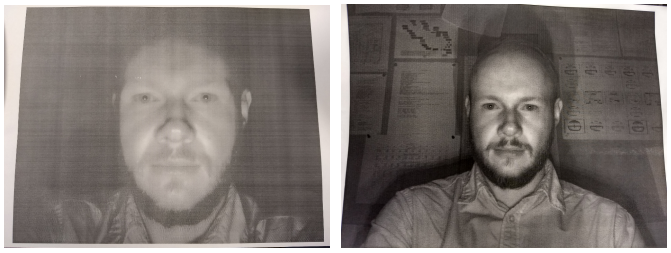
We did this for two sensors, NIR Fast and NIR Slow and two different users. We used three different printers, two gray scale laser printer and one color laser printer, from two manufacturers, HP and RICOH. The quality of the prints ranged from bad



(a) Setup of NIR image acquisition.



(b) Illumination of the biometric capturing device.



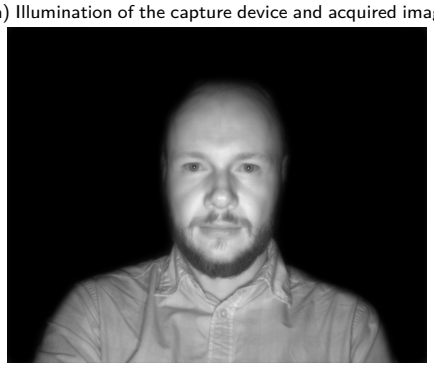
(c) Print of the captured image in various qualities, both were successfully used in presentation attacks.

Figure 3: A NIR image attack, capturing, artefacts and presentation attack.

to OK, Fig. 3c, but in each case the presentation attack/liveness detection could successfully be circumvented. The main hurdle for a successful attack was that the reflective glare from the paper could confuse the sensor, but a slight bending of the paper to prevent a direct reflection solved this in all cases. What this means is that the shift from the visible spectrum to the NIR spectrum was done purely to make image acquisition harder, not to allow for more sophisticated liveness detection methods. This move will likely prevent a casual attacker from succeeding, but even a moderately committed attacker can easily overcome this shift by simply shifting the image acquisition to NIR also.

However, we have cheated a bit by using the illumination of the sensor during the image acquisition. This might be hard to do in the wild and it creates an image specifically expected by the sensor given the illumination, making the attack a bit easier than it would be in the wild. To see if we could successfully mount an attack with a different illumination we used the image acquired from the NIR Slow, the sensor with the stronger illumination, and attempted to attack the NIR Fast system with it. Figure 4a shows the sensor illumination and acquired image. Initial attacks failed because the sensor miss-identified the background as part of the face. This is due to the low illumination of the NIR Fast achieving a good separation of subject and background. To

(a) Illumination of the capture device and acquired image.



(b) The image used in the attack.

Figure 4: A NIR presentation attack using a different illumination during capture and attack. The attack image used image manipulation after acquisition to separate the subject from the background.

fix this we used an image with a darkened background as PAD artefact, Fig. 4b. This succeeded in unlocking the device.

While not exhaustive, we have shown that the shift to NIR does not improve security by itself. Similar attacks as in the visible spectrum are possible.

Complexity of Attacks

The complexity of an attack in the NIR spectrum, as it currently stands, is in the realm of 'Level B' to 'Level C'. The specific knowledge required is not particularly difficult, but a certain amount of planning, money, and time is still required for this type of attack. The main problem is the acquisition of a NIR image which also requires a NIR camera. However, this does not mean that NIR imaging is a solution to current visible light attacks. If more COTS smartphones would contain a NIR imaging system, a prerequisite for NIR based recognition, then the main hurdle for the attack, the requirement of a NIR imaging device, is also removed!

How to test liveness detection without a NIR capturing device

Laser toner ink (we have not tested ink printers) absorbs NIR light. This means a printed face, even if not recorded in the NIR spectrum has a representation in the NIR spectrum. Therefore, a printed face can be enrolled and used for an unlock attempt. The unlock attempt is genuine since the same artefact is used for enrollment and for unlocking. The only system preventing this would be liveness detection. This method can be used for a low cost and low effort evaluation of the liveness detection presence and capabilities. As an example, for NIR Slow and NIR Fast this simple test allows to enter the system. This means that the only

liveness detection is the use of the NIR spectrum. This in essence is not a liveness detection at all, it simply complicates an attack a bit.

6 Usability

6.1 Time to unlock

We should mention that we are talking about unlocking a smartphone for immediate use us our use-case. This is a frequent task and the user expects a timely response. For use cases which are less frequent and thought off as requiring a higher security, e.g., banking applications, the willingness of users to wait longer in trade-off for an increase in security can be assumed. As mentioned in Section 4, we recorded the time to unlock (TTU) measuring user satisfaction by means of how long it takes to unlock a device. Under optimal conditions, i.e., natural light, all generic systems show rather long TTUs as illustrated in Table 5. The TTU is furthermore highly dependent on the used device. The fastest combination of system and device unlock after around 13 seconds, while the worst combinations take over 40 seconds or do not even unlock at all. Dedicated systems on the other hand perform much better due to the dedicated hardware and are able to achieve TTUs as low as 2.1 seconds for NIR Fast. The slowest dedicated system, Not NIR still achieves comparable performance to the best generic systems. These results clearly illustrate the advantage of using dedicated hardware, in this case a shift of the imaging from the visible to the NIR spectrum, to enhance user satisfaction and therefore usability.

6.2 Varying illumination

The tested variations in illumination (spot and diffused light from the front, back and side) caused major problems to all generic systems, which manifest themselves in form of failed verification attempts and increased TTU. The illumination changes interfered with the systems by messing with the face detection or decreasing the contrast of the acquired images in order to prevent processing of the images due to insufficient quality. Furthermore, the behavior of the systems differs among the tested smartphones, e.g., a system might work well with spot-light from the back with one smartphone, but not work at all with another smartphone. An overview of the most problematic illumination scenarios for the investigated systems is given in Table 7. We observed that spot light in general, and with high intensity in particular, as well as diffused light with low intensity caused the most problems for the generic systems. Spot light often causes reflections on the skin, which often trigger the liveness detection subsystem into falsely detecting a presentation attack, while the low diffused light leads to insufficient quality of the acquired images or videos. We observed, that also the employed lens systems play a significant role for the obtained image quality. The lens system has to be able to prevent flares and retain contrast in the image with varying illumination conditions, especially with light from the back.

Similar to the previous usability analysis of the TTU, the dedicated hardware offers big advantages over generic systems.

Table 7: Overview of most problematic illumination scenarios for generic systems. The illumination type is given by a combination of letters (**D**iffused, **S**pot) and digits (intensity) applied from different directions (**F**ront,**S**ide,**B**ack).

	O+ 6			SG S8			SXC 4			LG K8		
	F	S	B	F	S	B	F	S	B	F	S	B
LD only	D1	S6		S1			D1	S1				
FollowIt				S1	S6	S6				S1	S6	S6
OnDevice: Active					S6	S6				D1	S6	D6
OnDevice: Passive					S6	S6				D6	S6	S6
OnServer C4	S6	S1	S3	S6	D6	S6	D1	D6	S6			
OnServer C6	S6	S1	S6	D1	D1	D6	S1	S1	S6			

Especially the use of NIR imagery, i.e., dedicated camera and illumination as compared to generic systems that rely on the environment to illuminate the subject, makes the dedicated systems basically insensitive to illumination changes with exception of the Not NIR system.

7 Lessons Learned and Practical Tips

This section presents some lessons we have learned, which might benefit other researchers so that a they can learn from our mistakes. These tips are for presentation attack attempts, not genuine unlock attempts.

Mind lighting conditions, specifically make certain that (A) the parts of the biometric characteristic are well lit while (B) reducing strong reflections, glare and highlights as much as possible. Some examples of this are:

- Eyes have always to be well lit when using any mask, we looked up into the ceiling light while unlocking to ensure an even illumination, see Fig. 2c, right image.
- For replay attacks with screens, reduce smartphone screen and increase attack screen illumination as much as possible. This ensures that the screen of the smartphone is not reflected back into the presentation attack artifact.
- Similarly, for replay attacks with screens, distance from the camera and tilting of the screen to prevent direct reflections back at the camera are important.
- This also holds for other presentation attacks where the sensor includes an illumination module, like in the case of NIR sensors. In this case a bending of the paper to scatter the light was sufficient to circumvent problems.

Prints should have a high quality to increase the chances of the attack. The background in images can confuse the sensor, so either be mindful of the background or separate the subject and background artificially, e.g., Fig. 4b.

Be creative when using paper prints to circumvent interaction based liveness detection, e.g., Fig. 1. We found these kind of attack much easier to mount than video based attacks since we had more control over the interaction. **Note:** This is also an indicator that interaction based liveness detection is much less useful than we might think!

Table 8: A summary of the finding broken down into simple categories.

System	Devices	Security	Light Sensitivity	TTU	Overall
LD only	3/4	Bad	Bad to Passable	Bad	Bad
FollowIt	2/4	Bad	Bad	Bad	Bad
OnDevice: Active	2/4	Bad	Bad	Bad	Bad
OnDevice: Passive	2/4	Bad	Bad	Bad	Bad
OnServer: lenient	4/4	Bad	Bad to Passable	Bad	Bad
OnServer: strict	4/4	Bad	Bad	Bad	Bad
FaceID	–	Good	Good	Good	Good
Not NIR	–	Bad	Bad	Bad	Bad
NIR Slow	–	Passable	Good	Passable	Passable
NIR Fast	–	Passable	Good	Good	Passable

Devices: The number of mobile devices (smartphones) on which the system could be deployed (4 in total), only relevant for generic systems.
Security: *Bad* if breakable by a 'Level A' or 'Level B' attack, *Passable* if breakable by a 'Level C' attack. *Good* if not breakable (in our tests).
Light Sensitivity:*Bad* if natural light is less than 100%, *Passable* if Bad and can be unlocked in every light condition, *Good* if Passable and no light scenario below 75% chance of unlock.
TTU: *Bad* if more than 10s, *Passable* if between 5 and 10 seconds, otherwise *Good*[3]
Overall: Minimum of TTU, Light Sensitivity and Security.

8 Conclusion

The overall results with a simple categorization are given in Table 8. From this table it is pretty clear that creating a generic system is challenging. Currently we would classify each of the tested systems as unusable. The situation is better for systems with dedicated hardware. The FaceID is a perfectly usable system as far as we could tell. This shows that, even with limited space for dedicated hardware, a fast and reliable system can be built. However, dedicated hardware is indeed required, simple use of COTS hardware without sophisticated software is not enough, this is showcased by the stationary systems.

The main problem in most cases seems to be a poor translation of, even basic, techniques and methods from literature to commercial products. This problem specifically pertains to liveness detection, biometric comparison worked well in almost all cases. The main problem with liveness detection seems to be harsh lighting conditions which normally do not occur in lab environments.

Acknowledgements

This work has been partially supported by Veridos LLC.

[1] Unfortunately, there is no reference on this. We have set this to 5 seconds based on experience. This is, in our humble opinion, already far too long for unlocking a smart phone. Just think about this: What is an acceptable time to unlock the device if it *always* takes that time (we are talking average time to unlock, not worst case).

References

[1] 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)*, 2019. DOI: 10.1109/ICB45273.2019.8987404 (cit. on pp. 2–4).

[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 (cit. on p. 3).

[3] M. Chrzan, "Liveness detection for face recognition," Master's thesis, Masaryk University, Factory of Informatics, 2014 (cit. on p. 3).

[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 (cit. on p. 3).

[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 (cit. on p. 3).

[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 (cit. on p. 3).

[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 (cit. on p. 3).

[8] 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 (cit. on p. 3).

[9] 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 (cit. on p. 3).

[10] 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 (cit. on p. 3).

[11] D. Wen, H. Han, and A. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, 2015 (cit. on pp. 3, 4).

[12] I. Chingovska, N. Erdogmus, A. Anjos, and S. Marcel, "Face recognition systems under spoofing attacks," in *Face Recognition Across the Imaging Spectrum*, 2016 (cit. on pp. 3, 4, 10).

[13] 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 (cit. on p. 3).

[14] 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 (cit. on p. 3).

[15] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based countermeasures to photo attacks in face recognition," *IET Biometrics*, 3 2014 (cit. on p. 3).

[16] 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 (cit. on p. 3).

[17] Y. Li, Y. Li, K. Xu, Q. Yan, and R. H. Deng, "Empirical study of face authentication systems under OSNFD attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 2, 2018 (cit. on pp. 3, 4).

[18] C. Chen, A. Dantcheva, T. Swearingen, and A. Ross, "Spoofing faces using makeup: An investigative study," in *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2017 (cit. on pp. 3, 4).

[19] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore, "Face presentation attack with latex masks in multispectral videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017 (cit. on pp. 3, 4).

[20] R. Ramachandra, S. Venkatesh, K. B. Raja, S. Bhattacharjee, P. Wasnik, S. Marcel, and C. Busch, "Custom silicone face masks: Vulnerability of commercial face recognition systems & presentation attack detection," in *2019 7th International Workshop on Biometrics and Forensics (IWBF)*, IEEE, 2019 (cit. on pp. 3, 4).

[21] S. Jia, G. Guo, Z. Xu, and Q. Wang, "Face presentation attack detection in mobile scenarios: A comprehensive evaluation," *Image and Vision Computing*, vol. 93, 2020, ISSN: 0262-8856. DOI: 10.1016/j.imavis.2019.11.004 (cit. on pp. 3, 4).

[22] S. Jia, G. Guo, and Z. Xu, "A survey on 3d mask presentation attack detection and countermeasures," *Pattern Recognition*, vol. 98, 2020, ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2019.107032 (cit. on pp. 3, 4).

[23] R. Raghavendra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Computing Surveys*, vol. 50, no. 1, 2017 (cit. on p. 2).

[24] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015 (cit. on p. 4).

[25] S. Schuckers, "Presentations and attacks, and spoofs, oh my," *Image and Vision Computing*, vol. 55, 2016, Recognizing future hot topics and hard problems in biometrics research. DOI: 10.1016/j.imavis.2016.03.016 (cit. on p. 5).

[26] ISO/IEC 30107-3, *Information technology - biometric presentation attack detection - part 3: Testing and reporting*, 2017 (cit. on p. 6).

[27] L. Debiasi, C. Kauba, H. Hofbauer, B. Prommegger, and A. Uhl, "Presentation attacks and detection in finger- and hand-vein recognition," in *Proceedings of the Joint Austrian Computer Vision and Robotics Workshop (ACVRW'20)*, 2020. DOI: 10.3217/978-3-85125-752-6-16 (cit. on p. 10).

[28] P. Tome and S. Marcel, "On the vulnerability of palm vein recognition to spoofing attacks," in *The 8th IAPR International Conference on Biometrics (ICB)*, 2015 (cit. on p. 10).