Recompression Effects in Iris Recognition[☆]

Thomas Bergmüller[a,b], Eleftherios Christopoulos[b], Kevin Fehrenbach[b], Martin Schnöll[b], Andreas Uhl[b,*]

[a]*Authentic Vision GmbH, Austria*
[b]*Department of Computer Sciences, University of Salzburg, Austria*

Abstract

Rating a compression algorithms' performance is usually done in experimental studies, where researchers have frequently used JPEG pre-compressed data. It is not clear yet, if results of such compression experiments are reliable when conducted on pre-compressed data. To investigate this issue, we first study the impact of using pre-compressed data in iris segmentation and evaluate the relation between iris segmentation performance and general image quality metrics. In this context we propose a method to overcome potential problems in case using pre-compressed data sets cannot be avoided. As the second step, we conduct experimentation on the entire iris recognition pipeline. We find that overall, recognition accuracy results might not be entirely reliable in case of applying JPEG XR or JPEG2000 to JPEG pre-compressed data.

1. Introduction

Iris recognition [1, 2] is one of the most deployed biometric modalities, standardised by the International Civil Aviation Organisation (ICAO) for use in future passports, and one of the technologies in the Unique Identification Authority of India's (UID) Aadhaar project to uniquely identify Indian citizens. The increasing market saturation of biometrics instead of conventional access control methods raises the need for efficient means to store such data. The International Organisation for Standardisation (ISO) specifies iris biometric data to be recorded and stored in (raw) image form (ISO/IEC FDIS 19794-6) rather than in extracted templates (e.g. iris-codes). Such deployments benefit from future improvements (e.g. in feature extraction stage) which can be easily incorporated without re-enrollment of registered users. Since biometric templates may depend on patent-registered algorithms, databases of raw images also enable more interoperability and vendor neutrality [2]. These facts motivate detailed investigations and optimisations of image compression on iris biometrics in order to provide an efficient storage and rapid transmission of raw biometric records. Furthermore, the application of low-powered mobile sensors for image acquisition, e.g. mobile phones, raises the need for reducing the amount of transmitted data.

As a consequence, according to the importance of this issue, many studies comparing and optimising lossy compression techniques for iris imagery may be found in literature. Since the CASIA iris datasets have been very popular among researchers ever since their establishment, many papers dealing with compression have been relying on the (extended) CASIA V1.0 dataset, including also first IREX investigations [3, 4, 5, 6, 7] (apart from other examples using the ICE 2005 dataset [8, 9]).

Since it has been pointed out [10] that the CASIA V1.0 dataset exhibits manipulated pupil areas and should therefore not be used any further in experimentation, compression researchers moved to other (and more recent, more challenging etc.) datasets, e.g. the CASIA V3.0 [11, 2], the CASIA V4.0 [12], the Bath [4, 13], and the UBIRIS.v1 [6, 14] datasets. While the images of CASIA V1.0 and ICE 2005 are given in uncompressed format, images in CASIA V3.0, CASIA V4.0, UBIRIS and Bath datasets are provided as JPEG (the first three) or JPEG2000 (the latter) lossy compressed data. Therefore, any compression experiments conducted on these datasets operate on pre-compressed data. This fact has not been ignored entirely – for example, in [2], preparatory JPEG compression experiments with uncompressed data reveal that slightly pre-compressed data leads to better recognition performance due to denoising effects. Thus experiments with pre-compressed data are assessed to be unproblematic. The same argument is used for JPEG2000 pre-compressed data [13]: In [4] it was also shown that slight pre-compression with JPEG2000 improves recognition rates, thus JPEG2000 pre-compression is not seen problematic in any way. However, eventual artifacts resulting from recompression effects are not accounted for in these considerations. Recompression artifacts arise in cases where data is compressed twice (or multiple times) with lossy compression schemes, i.e. where artifacts from the first compression step (termed pre-compression) are aggravated or exploited by the second compression step.

Two different types of such effects may be distinguished: First, *homogeneous recompression*, where the same compression scheme is used several times, whereas in *heterogeneous recompression* different methods are used in the different compression steps. For example, in iris recognition, using JPEG pre-compressed data and applying JPEG XR and JPEG2000

[☆]This work has been partially supported by the Austrian Science Fund FWF, project no. P26630.

[*]Corresponding author
Email addresses: tb@authenticvision.com (Thomas Bergmüller), uhl@cosy.sbg.ac.at (Andreas Uhl)

[11] or JPEG2000 and fractal compression [14] is eventually prone to heterogeneous recompression artifacts, while the application of JPEG to JPEG pre-compressed data [12, 2] can be prone to homogeneous recompression artifacts. However, experiments prone to eventual recompression effects are not limited to compression in iris recognition. As a further example we consider compression in face recognition, where [15] uses the CMU Face In Action (FIA) Database (which is in JPEG format) to investigate H.264 compression effects and [16] eventually suffers from homogeneous recompression effects since the original images shown in the paper before applying JPEG compression clearly already exhibit JPEG compression artifacts.

While next to nothing can be found on the issue of heterogeneous recompression artifacts in the general compression literature, homogeneous recompression artifacts are better investigated, at least in the case of lossy JPEG compression. Soon after the establishment of the JPEG standard [17], it was found that JPEG recompression artifacts arise and do not follow a linear behaviour [18]. Extensive experiments in this direction can also be found in [19], and following these observations, requantisation-based schemes have been suggested for JPEG, reducing recompression artifacts considerably [20]. Recently, the identification of images which underwent JPEG double compression (i.e. JPEG homogeneous recompression) has been a hot topic in image forensics [21]. Taking all these facts together, it gets clear that recompression artifacts may impact experimental results with respect to biometric recognition performance, an issue, that has been neglected so far. As discussed, ISO/IEC FDIS 19794-6 requires storing biometric data as raw images, hence all components of a biometric system are affected when operating with compressed data. As we investigate the recompression issue by studying impact on an iris recognition system, the influence on segmentation and texture extraction as well as feature extraction, i.e. iris code computation, has to be evaluated. [22, 8] suggest that data reduction has the highest impact on the iris segmentation. Since segmentation is also the first step in the pipeline, this potentially effects the performance of later steps as well and is therefore of particular importance.

We systematically investigate eventual homogeneous and heterogeneous recompression effects in an experimental study for iris recognition. Given the importance of JPEG (as the CASIA V3.0/V4.0 and UBIRIS.v1 datasets are only available in this format), we focus on JPEG pre-compressed data. In our experiments, we first compare iris segmentation and general purpose image quality metrics applied to single compressed vs. recompressed (i.e. JPEG pre-compressed) iris image data, before assessing eventual recompression effects on iris recognition accuracy. Section 2 discusses relevant aspects on generating single- and recompressed data. The used data sets and methods are described in section 3. Section 4 introduces several quality metrics and iris segmentation experiments and lists their results, which are then compared in section 5. Section 6 presents experimental results on iris recognition accuracy in recompression scenarios. From the experiments' individual and comparison results, we draw conclusions in section 7.

2. Compression scheme

As discussed, we investigate whether there is a difference in using truly uncompressed data or pre-compressed data in experiments rating the performance of an iris segmentation. Using pre-compressed data means a pre-compressed image I_p is compressed a second time, resulting in an recompressed image I_r. When compressing a truly uncompressed image I_u, the resulting image I_s is generated in a single compression step. Since experiments are typically carried out on a data set with more than one image, we denote $I_u^{(k)}, I_p^{(k)}, I_s^{(k)}, I_r^{(k)} \in \mathbb{R}^{w \times h}$ as the k^{th} image with width w and height h. For simplicity, I_u, I_p, I_r, I_s subsequently denote a particular but unspecified image of a data set. Furthermore, we define $s(F) \in \mathbb{N}$ with $F \sim I \in \mathbb{R}^{w \times h}$ as a function that returns the file size of the file F storing an image I. Since common lossy compression algorithms also employ lossless compression methods, e.g. run-length encoding, before writing to a file, F is only loosely related to the pixel data, namely the image, I. For simplicity, we denote $s(I)$ as the file size of the file F encoding the pixel values of an image I. $c_m(I, q)$ with $q \in \mathbb{N}$ describes the process of compressing an image I using a particular method m parametrised with the quality parameter q. In terms of this paper we use the values $m \in \{jpg, jxr, j2k\}$, where

- jpg corresponds to the well-known (ISO/IEC IS 10918-1) DCT-based image compression method JPEG,
- $j2k$ corresponds to the wavelet-based image compression standard JPEG2000 (ISO/IEC IS 15444-1), which can operate at higher compression ratios and
- jxr corresponds to the compression standard JPEG-XR based on Microsofts HD Photo, which is specified in (ISO/IEC IS 29199-2).

For rating an image I's compression effectiveness, we define the compression ratio cr between an uncompressed image I_u and a compressed image I_c as

$$cr(I_u, I_c) = \frac{s(I_u)}{s(I_c)} \quad \text{with} \quad I_c \in \{I_r, I_s\} \tag{1}$$

For the later described experiments, images are compressed to a target compression ratio $cr_t \in \mathbb{R}$. However, only the *j2k* compression standard allows to specify a target compression ratio cr_t directly via parameter q. Hence this is the only method where we can control the file size $s(I_c)$ directly. The other two compression methods take a quality parameter $q \in \mathbb{N}$ only, controlling the quality but not the file size $s(I_c)$. Thus it is not possible to set this parameter to meet a certain target compression ratio cr_t. Due to the quality parameter's limited set of quality values, the target compression ratio cr_t cannot be achieved exactly for any of the three methods. Parameter optimisation can be done, such that $cr_t \approxeq cr(I_u^{(k)}, I_c^{(k)})$. We propose an algorithm to compress a set of K uncompressed images I_u using a particular method m to achieve a certain compression ratio cr_t in a way that the compression ratio of each image is met as closely as possible. This process, illustrated in Fig. 1, employs

1. Compute the single-compressed image $I_s^{(k)}$ with method m such that $cr(I_u^{(k)}, I_s^{(k)}) \approx cr_t$. The optimal quality parameter $q_s^{(k)}$ is computed for each image separately by

$$s_t^{(k)} = \frac{s(I_u^{(k)})}{cr_t} \tag{2}$$

$$q_s^{(k)} = \underset{q \in \mathbb{N}}{argmin} |s(c_m(I_u^{(k)}, q)) - s_t^{(k)}|, \tag{3}$$

where $s_t^{(k)}$ is the file size exactly meeting the target compression ratio cr_t. This is implemented by iteratively searching the quality parameter q that results in the closest achievable compression ratio $cr(I_u, I_c)$. The single compressed images $I_s^{(k)}$ using method m are computed with the optimal parameters $q_s^{(k)}$ as

$$I_s^{(k)} = c_m(I_u^{(k)}, q_s^{(k)}) \tag{4}$$

2. Compute a pre-compressed image $I_p^{(k)}$ using jpg-method with an arbitrary but fixed quality parameter q_p, i.e.

$$I_p^{(k)} = c_{jpg}(I_u^{(k)}, q_p) \tag{5}$$

3. Now, find a quality parameter $q_d^{(k)}$ that allows to compress the pre-compressed image $I_p^{(k)}$ a second time, such that the resulting recompressed image $I_r^{(k)}$ has the same file size as the single-compressed image $I_s^{(k)}$, i.e. $s(I_s^{(k)}) \approxeq s(I_r^{(k)})$. Such a quality parameter $q_r^{(k)}$ can be found by optimising

$$q_r^{(k)} = \underset{q \in \mathbb{N}}{argmin} |s(c_m(I_p^{(k)}, q)) - s(I_s^{(k)})| \tag{6}$$

$$\forall s(I_s^{(k)}) \geq s(I_r^{(k)}) \tag{7}$$

The condition $s(I_s^{(k)}) \geq s(I_r^{(k)})$ is of importance to establish fair conditions, since it is very likely that the file sizes $s(I_s^{(k)}), s(I_r^{(k)})$ cannot be equalised due to the limited set of the quality parameters q. The recompressed images $I_r^{(k)}$ are then computed from the pre-compressed images $I_p^{(k)}$ with the found optimal parameters q_r as

$$I_r^{(k)} = c_m(I_p^{(k)}, q_r^{(k)}) \tag{8}$$

Using data sets generated with this method, we can investigate the impact of artifacts in an recompressed image I_r in comparison to those in single-images I_s. The first one contains artifacts by two compression stages, while the latter one contains artifacts from one compression step only.

3. Experimental setup

Although there are several iris data sets around, few are available in uncompressed format. We use the IITD Iris data base[1]. The main reason for this is the availability of a segmentation ground truth created by an expert, which was recently

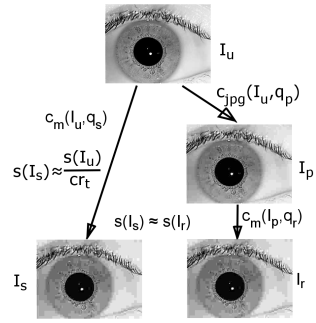
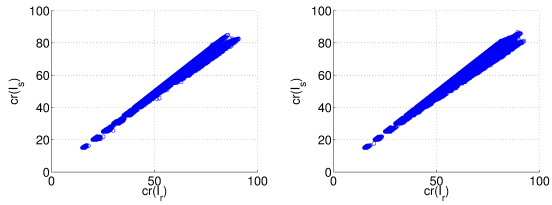


Figure 1: Compression principle to obtain two images achieving approximately the same target compression ratio cr_t from an uncompressed image $I_u^{(k)}$ using a particular compression method m. One image, $I_s^{(k)}$, is compressed in a single step while the other, $I_r^{(k)}$, uses a pre-compression and a final compression step. The pre-compression step is always a jpg-compression, while the final one uses the same method m as used in single-compression.

introduced by Hofbauer et al. [23] and used in [22]. The k^{th} image of this segmentation ground truth data set is subsequently denoted as $SGT^{(k)}$. According to information by the IITD iris data base's authors, the images, stored in a 3-channel uncompressed bitmap format[2], are already JPEG-compressed with 100% quality by the sensor (JIRIS, JPC1000). Since they are stored as bitmaps, all images have an identical file size of $s(I_u)$=230,454 bytes. While this reference file size is used for experiments on image quality metrics and segmentation, we use the "correct" original $s(I_u)$=76800 bytes filesize in the context of iris recognition to consider even stronger compression (since the only significant differences between single compression and recompression in iris segmentation have been found for high compression ratios [24]). Thus, when comparing identical compression ratios between those two sections, they actually differ by a factor of 3. Despite not being optimal, using the IITD was necessary due to the available ground truth, for reasons becoming obvious in section 4.3. Furthermore, the IITD – contrary to others, e.g. the ND-IRIS-0405 iris image dataset [25] – is captured under favorable conditions, which allows for lower segmentation errors. This is necessary to distinguish between noise and recompression-effects. We use the scheme introduced in section 2 to compress obtain data sets with target compression ratios $cr_t \in \{15, 20, 25, ..., 70, 75\}$. For each of these target compression ratios cr_t, the pre-compression step in recompression mode is carried out with quality parameters $q_p \in \{100, 80, 75, 70\}$ to simulate different levels of pre-compression. Each of these combinations is used to compress with the introduced jpg, $j2k$ and jxr methods. We start at compression ratio $cr_t = 15$, because even a pre-compression with $q_p = 100$ achieves - depending on the image's content - already a compression ratio of $cr(I_u, I_p) \approx 10$. For obvious reasons, no smaller compression ratio $cr(I_u, I_r) < cr(I_u, I_p)$

[1] IITD Iris Database version 1.0, `www4.comp.polyu.edu.hk/\textasciitildecsajaykr/IITD/Database\_Iris.htm`

[2] We want to point out that storing in 1-channel bitmaps as done for recognition experiments would be more efficient, since the images were captured in near-infrared. However, we use the size information of the 3-channel bitmap in computing compression ratios



	single: $1 - \frac{cr(I_u,I_s)}{cr_t}$			recomp.: $1 - \frac{cr(I_u,I_r)}{cr_t}$		
%	*jpg*	*j2k*	*jxr*	*jpg*	*j2k*	*jxr*
μ	-3.21	-3.48	-4.33	-6.80	-7.04	-9.49
σ	2.74	2.51	2.87	4.83	4.23	4.34

Figure 2: Scatter plots of measured compression ratios $cr(I_u, I_s)$ over $cr(I_u, I_r)$ for methods *jpg* (left) and *jxr* (right). The graphs indicate that the $s(I_s^{(k)}) \geq s(I_r^{(k)})$ condition from equ. (7) is satisfied. While this is indeed true for *j2k* and *jxr*, we observe a violation in 0.13% of the cases for *jpg* at $cr_t \geq 70$, because JPEG is already operating at its bitrate limits at such high compression ratios. The table below the figures reveals, that on average the aimed cr_t is met with 3.67% accuracy for single-compressed images I_s, while the recompressed ones I_r only reach 7.8%. This is due to the limited set of quality parameters q, q_p.

can be reached in recompression. This results in a total of 195 data sets with 2240 images each, whose distribution is shown and discussed in Fig. 2.

4. Evaluation

We investigate the behaviour of iris segmentation employing a segmentation error rate (section 4.3). Besides that, we assess the image quality with full-reference metrics (section 4.1) as well as non-reference metrics (section 4.2). The individual results are then compared in section 5.

As correctly pointed out [26], image quality is by no means equivalent to biometric quality. For example, an image of the eye with closed lid might exhibit excellent image quality while it is of lowest biometric quality in an iris recognition context. However, images with low image quality will always also have reduced biometric quality, as shown for many modalities e.g. when samples are compressed. Therefore, it makes sense to apply image quality metrics to look for eventual effects of recompression in order to be able to identify occurring effects by just evaluating such metrics on the data instead of being forced to conduct the more costly segmentation or even recognition experimentation.

4.1. Full-reference quality metrics

Evaluating the quality of the compressed images with respect to the original an assortment of full-reference metrics was chosen. The choice was made according to different aspects of human perception starting from mathematically defined to low-level features based and finally to high-level features based. It has to be pointed out that in general, the application of full-reference metrics in a biometric environment is difficult as in most cases, no undistorted original is available for the assessment. However, in recent work it has been shown that the usage of a different image of the subject to be authenticated (eventually stored from previous enrollment) can be used for quality assessment with full-reference quality metrics quite successfully [27]. The following metrics were included:

- PSNR: Peak signal-to-noise ratio.
- MSSIM [28]: Multi-scale structural similarity index is an extension of the SSIM metric. After the extraction of luminance, structure and contrast components from the image at scale 1, the algorithm iteratively applies a low pass filter and downsamples the filtered image by a factor of 2. The overall result is the combination of measurements at different scales.
- NQM [29]: Noise Quality Measure, a low-level HVS features based metric. The contrast pyramid of Pelis work was used to model the variation in contrast, sensitivity with distance, dimensions and spatial frequency of the stimuli, and with the variation of their local luminance mean.
- RFSIM [30]: Riesz-transform feature based similarity metric approximates HVS by perceiving an image mainly according to its low-level features and uses the 1st-order and 2nd-order Riesz transform coefficients. The similarity index is measured by comparing the two feature maps at key locations marked by the feature mask. The mask is generated by a Canny operator.
- VSNR [31]: Visual Signal-to-Noise Ratio, a wavelet based metric. The metric is designed to evaluate both low-level and mid-level HVS features. VSNR works in two stages: The first computes the contrast detection thresholds, while the second estimates visual fidelity by measuring the perceived contrast and the extent to which the distortions disrupt global precedence.

Applying these quality metrics, *jpg*, *j2k* and *jxr* resulted in the following observations (see Fig. 3 and 4):

1. For *jpg* and $cr_t > 15$ single compressed images were of higher quality compared to recompressed ones and at cr_t=15 the single compressed images were of the lowest quality as shown in Fig. 3 for MSSIM. The quality of the recompressed images followed the trend that the higher the q_p the better the quality of the image. The previous observation of metrics for *jpg* was unanimous for all metrics.
2. For *jk2* in all compression ratios and all metrics, the quality of the recompressed images followed the trend that the higher the q_p the better the quality of the image. Single compressed images were of the highest quality compared to recompressed data, which was valid for all metrics and compression ratios as shown exemplary for MSSIM in Fig. 3.
3. For *jxr* and $15 \leq cr_t \leq 40$ single compressed images were of the lowest quality compared to recompressed data for MSSIM and VSNR. The latter followed the trend of the higher the q_p the better the quality of the image (Fig.

4). For $45 \leq cr_t \leq 75$ images of single compression became of the highest quality and recompressed data continued the same trend for MSSIM and VSNR metrics. NQM showed the same behaviour, but cr_t=50 was observed to be the changing point in this case. RFSIM showed a different trend from the previous; single compressed data were always of the best quality compared to recompressed data which followed in terms of quality measurement.

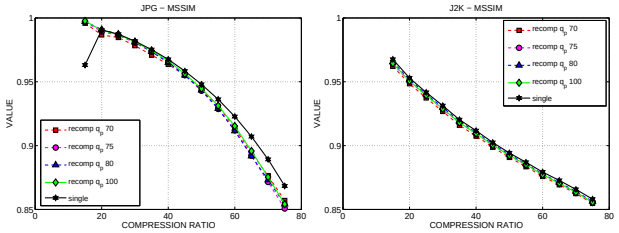


Figure 3: Left: MSSIM of *jpg* single- and recompressed data. Right: MSSIM of *j2k* single- and recompressed data

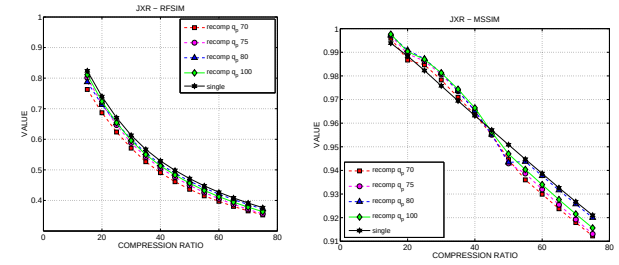


Figure 4: Left: RFSIM of *jxr* single- and recompressed data. Right: MSSIM of *jxr* single- and recompressed data

4.2. Non-reference quality metrics

Out of the four metrics considered in this work, NRPQA and BRISQUE have been already employed in a biometric context recently [26].

- ANISOTROPY [32]: The anisotropy is measured as the difference from the expectation value of the entropy for different orientations in the image (where pointwise Renyi entropy is actually computed). A distinct maximum value is obtained for undistorted images.

- NRPQA [33]: Primarily measures compression artifacts resulting from block-based lossy compression schemes (like JPEG and JPEG XR). It is computed in the spatial domain as the combination of blockyness and activity estimation (the latter reduced by removal of high frequency information) in both horizontal and vertical directions, manifesting in three metrics: Blockyness, activity, and zero-crossing rate. As we apply JPEG-based pre-compression, all types of double compression schemes could / should be detected with this metric.

- BRISQUE [34]: This quality metric is a deviation measure of a natural image from the regular statistics, indicating distortion. It may thus be interpreted as a holistic assessment of image naturalness. The underlying features are computed from the empirical distribution of locally normalised luminances and products of locally normalised luminances under a spatial natural scene statistic model.

- NIQE [35]: NIQE is based as well on a quality aware collection of statistical features based on a simple and successful spatial domain natural scene model. Contrasting to BRISQUE (and most other non-reference quality metrics), it is not trained on human-rated distorted images but the natural scene statistics are derived only from a corpus of natural, undistorted images without any exposure to distorted images.

Applying these quality metrics, *jpg*, *j2k* and *jxr* resulted in the following observations (see Figs. 5, 6 and 7):

1. For *jpg* we find the only two cases (shown in Fig. 5) where metrics behave as expected from the full-reference metrics' results. For increasing cr_t the superiority of the single compressed case also increases, q_{100} is second best, and the remaining q_p are indistinguishable (for BRISQUE and NRPQA). For ANISOTROPY, results are completely useless (compare also Fig. 7), and for NIQE the expected behaviour is observed only for $cr_t \geq 50$ while for $cr_t \leq 30$, recompression q_{70} delivers the best values (ANISOTROPY and NIQE results not shown).

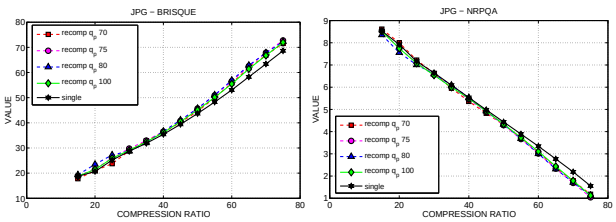


Figure 5: Left: BRISQUE of *jpg* single- and recompressed data. Right: NRPQA of *jpg* single- and recompressed data

2. For *j2k*, we observe two metrics where single compression is the best variant, and all other recompression settings perform worse, except for recompression q_{70} which is equally well than applying single compression (BRISQUE and NIQE, not shown). For $cr_t \geq 50$, for both NRPQA and ANISOTROPY at least the single compression case gives the best result, while the ranking of the other recompression cases is somehow unexpected (e.g. q_{75} worst and q_{70} best for ANISOTROPY). However, for $cr_t \leq 50$, for both metrics q_{70} results in the best overall values, q_{80} in the worst values, and the single compressed case is in-between (see Figs. 6 and 7).
3. For *jxr*, again the ANISOTROPY results are meaningless since the metric behaves reverse to its definition (metric values are better for higher cr_t, see Fig. 7) which is also the case for *jpg* and makes further discussion on results

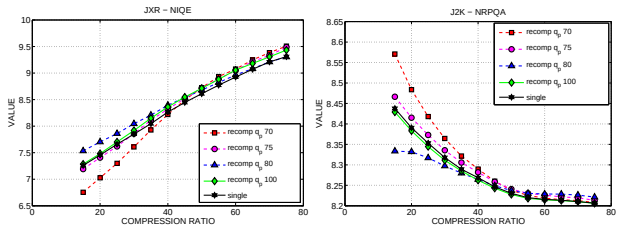


Figure 6: Left: NIQE of *jxr* single- and recompressed data. Right: NRPQA of *j2k* single- and recompressed data

meaningless. For the other three metrics, we observe a similar picture: For $cr_t \leq 50$, q_{70} results in the best overall values, q_{80} in the worst values, and the single compressed case is in-between (see Fig. 6 for the NIQE result). For $cr_t \geq 50$, the single compression case catches up to q_{70}. Note that this is quite similar to the results found for *j2k* and NRPQA as well as ANISOTROPY.

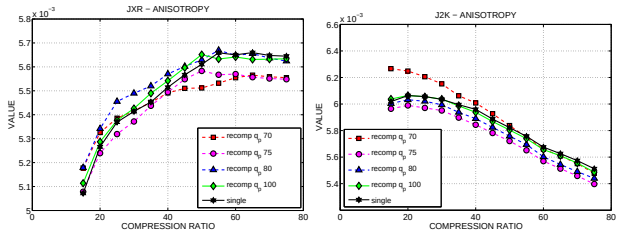


Figure 7: Left: ANISOTROPY of *jxr* single- and recompressed data. Right: ANISOTROPY of *j2k* single- and recompressed data

Overall, we find a much more non-uniform and inconsistent behaviour for non-reference metrics as compared to full-reference ones. However, still there are several cases (and also cr_t ranges), in which single compressed data is rated to deliver best quality, while especially in the range $cr_t \leq 50$ single compression is often found to be rated between q_{70} (being best) and q_{80} (being worst). ANISOTROPY completely fails to deliver useful results for *jpg* and *jxr*.

4.3. Segmentation error rates

In iris recognition, the segmentation of an iris image is considered as one of the most critical parts [8, 22]. We investigate the differences of single- and recompression as well as the aspects of which reference to use. We distinguish between using an absolute reference, e.g. a ground truth, and a relative one, e.g. the segmentation of the uncompressed images I_u, when computing the error rate.

The segmentation accuracy is rated by the mean segmentation error rate, which corresponds to the suggested E1 error rate in the Noisy Iris Challenge Evaluation - Part I (NICE.I). We define the segmentation error rate ser as

$$ser(R,S) = \overline{R \oplus S} \in [0,1] \quad with \quad R,S \in \{0,1\}^{w \times h}, \quad (9)$$

where R is the binarised reference segmentation and S the binarised segmentation result of the same image I. The mean

Figure 8: Segmentation masks of the expert ground truth [23], relative ground-truth $seg(I_u^{(k)})$ and an actual segmentation result $seg(I_r^{(k)})$ (f.l.t.r)

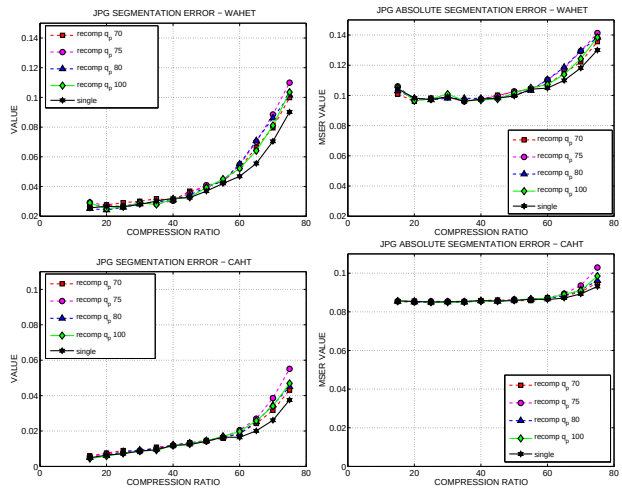


Figure 9: Relative and absolute segmentation error $mser_{rel}$ (left) and $mser_{abs}$ (right) with WAHET (top) and CAHT (bottom) segmentation on *jpg*-compressed data. Note that the $mser_{abs}$ is generally higher than $mser_{rel}$, because the tested algorithms ignore eyelids, yet they are considered in the expert ground truth [22].

value of the pixel-wise exclusive-or is the percentage of pixels different in the segmented image S in respect to the reference R. Due to multiple images in a data base, the mean segmentation error is computed from K images. We compute the absolute mean segmentation error $mser_{abs}$ in respect to the ground truth SGT and the relative mean segmentation error $mser_{rel}$ in respect to the segmentation of the uncompressed images I_u for single- and recompressed images $I_c \in I_s, I_r$. By denoting the segmentation result of an image I as $seg(I) \in \{0,1\}^{w \times h}$ we have

$$mser_{abs} = \frac{1}{K} \sum_{k=1}^{K} ser(SGT^{(k)}, seg(I_c^{(k)})) \quad (10)$$

$$mser_{rel} = \frac{1}{K} \sum_{k=1}^{K} ser(seg(I_u^{(k)}), seg(I_c^{(k)})) \quad (11)$$

The absolute segmentation error rate is considered to be optimal because of the available ground truth. However, for most data bases no such ground truth is available. Therefore we evaluate if the same conclusions as from the $mser_{abs}$ can be drawn from the $mser_{rel}$. The benefit of such a relation (if it exists) is that the $mser_{rel}$ can be computed for any arbitrary data set.

The data set described in section 3 is used to test the two

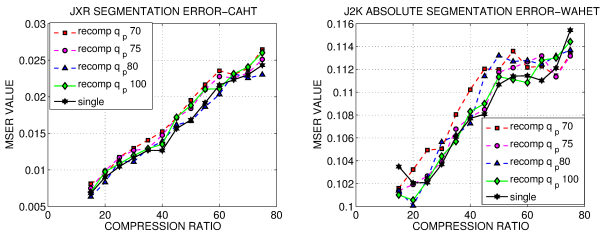


Figure 10: Relative CAHT segmentation error rate $mser_{rel}$ for *jxr*-compressed data (left) and absolute WAHET segmentation error rate $mser_{abs}$ for *j2k*-compressed data (right)

iris segmentation algorithms, Contrast-adjusted Hough Transform (CAHT) and Weighted Adaptive Hough and Ellipsopolar Transform (WAHET), from the USIT Framework v1.0.3[3]. The behaviour of these algorithms wrt. compression and impact of other artifacts is already analysed in literature [22, 2]. From our results we make the following observations:

1. Results for homogeneous recompression experiments, namely *jpg* on *jpg* pre-compressed data, in Fig. 9 indicate:
   (a) For small and medium compression ratios ($cr_t \leq 50$) no significant difference in segmentation errors of single- and recompressed data is observable. This implies that for these compression ratios it has no impact whether pre-compressed or uncompressed data is used in experiments.
   (b) For large compression ratios ($cr_t > 50$), segmentation errors tend to be lower for single-compressed data compared to recompressed data. Thus using pre-compressed or uncompressed data in experiments matters.
   (c) $mser_{rel}$ and $mser_{abs}$ generally show similar trends for medium and large compression ratios, i.e. there is a strong correlation of $mser_{rel}$ and $mser_{abs}$ for $cr_t > 30$. This means, the relative error $mser_{rel}$ suffices to rate performance on iris segmentation here, implying no expert-generated ground truth is needed.
   (d) However, WAHET segmentation errors reveal that in some cases there can be a difference between $mser_{rel}$ and $mser_{abs}$ for $cr_t \leq 30$. Fig. 9 shows here a different behaviour between absolute error $mser_{abs}$ (top-right) and relative error $mser_{rel}$ (top-left). Hence, for low compression ratios a ground truth is required.
   (e) In recompression, one might expect a linear relation between used pre-compression quality q_p and ranking of the error rates. Interestingly, when looking at Fig. 9 at high compression ratios, the poorest performance corresponds to $q_p = 75$, while the best is related to $q_p = 70$. In contradiction, $q_p = 100$ performs significantly better than $q_p = 80$ in most settings.

[3] as available at `http://wavelab.at/sources/` [2]

2. There are no clear trends for heterogeneous recompression experiments, namely *jxr* or *j2k* on *jpg* pre-compressed data. Even so, some interesting observations are made, which are illustrated in Fig. 10:
   (a) Data generated in a single compression step generally tends to result in smaller error rates compared to those computed from recompressed data. Interestingly, for extreme values, namely very small and very large compression ratios, single-compression performs often poorer.
   (b) For all experiments carried out with *jxr* and *j2k*, the error rate flattens in some way for medium compression ratios, i.e. $45 \leq cr_t \leq 70$. As an example this can be seen in the $mser_{abs}$ for *j2k* (Fig. 10 right). The characteristics of a curve's flattening vary depending on the pre-compression quality q_p. Since flattening can be seen in single- as well as recompressed data, we conclude the effect is generally related to the used methods *jxr* and *j2k*. However, the characteristics of the flattening seem to be controlled by the pre-compression quality q_p in a way that the lower the pre-compression quality is, the clearer the curve stagnates. Similar effects have been observed in a recent paper [36] where it turns out that strong compression artifacts are able to guide segmentation algorithms as those artifacts usually appear close to boundaries between differently structured areas (i.e. pupil, iris texture, sclera). Thus, the segmentation errors decrease while the compression ratio increases which can contribute to the observed flattening effect.

Thus, we reach the following conclusions:

(jpg) Homogeneous recompression. Recompression effects have a strong impact on experimental results for large compression ratios, i.e. $cr_t > 50$ (observations (1a),(1b)). Researchers are often forced to use pre-compressed data sets for the sake of ground truth availability. Results for compression ratios of $cr_t > 50$ can therefore not be considered entirely reliable (observation (1b)). However, recompression effects have negative influence on segmentation error rates, hence by using uncompressed data for the same experiments, better results may be achieved. If this behaviour is related to homogeneous recompression in general or for *jpg*-recompression only, is topic to further research.

From observation (1c) we know that for large compression ratios, i.e. $cr_t > 50$, there is no difference in the progress of $mser_{rel}$ and $mser_{abs}$. Since this is (from observation (1b)) exactly the range, where using single- or recompressed data does have an impact, we propose - based on observations (1d) and (1c) - to bench-mark compression algorithms in respect to iris segmentation by using

- uncompressed data sets rated with relative measures, such as the $mser_{rel}$, for severe compression, i.e. $cr_t > 50$ and

- pre-compressed data sets[4] with absolute measures, such as $mser_{abs}$, for medium and light compression, i.e. $cr_t \leq 50$.

If this applies to homogeneous recompression with other methods as well needs further investigation.

Heterogeneous recompression. We cannot observe comparable behaviour. However, there are trends observable (see observations (2a),(2b)), which need further investigation.

5. Comparison of Quality Metrics and Segmentation Accuracy

Besides evaluating the segmentation error rate and the general purpose quality measures independently, their correlation is analysed additionally. Furthermore, we explicitly investigate the correlation between $mser_{rel}$ and $mser_{abs}$ to back up the observations in section 4.3. For this purpose, we used the Spearman's rank correlation coefficient (SRCC) in two ways: First, computing quality metrics and segmentation errors per compression ratio averaged over all images and then correlating the two ordered arrays (SRCC$_A$, using the original definitions of $mser_{rel}$ and $mser_{abs}$), and second, computing SRCC for each image individually between the images' quality metric values and segmentation error per compression ratio and subsequently averaging the individual SRCCs (SRCC$_I$). For SRCC$_A$ we expect to learn about overall trends and more stable results as smaller inconsistencies wrt. single images are better hidden, however, it is not possible to provide confidence intervals using this approach. Thus, SRCC$_I$ is additionally considered to account more for the behaviour of single images (i.e. if it is possible to predict segmentation errors by computing metric values of single images).

In general, all five used general purpose full-reference quality metrics show a high linear relationship with a minimum SRCC$_A$ of 0.852 to each other. Furthermore, all of these metrics are highly correlating with the segmentation error rate and there are only minor differences between the segmentation algorithms (CAHT exhibits slightly higher SRCC$_A$ values than WAHET) and the relative and absolute segmentation error rates. Table 1 shows the SRCC$_A$ correlation results for the homogeneous recompression (*jpg*) and the CAHT segmentation algorithm. It can be observed that the $mser_{rel}$ shows overall a higher linear relationship to the quality metrics than the $mser_{abs}$. The reason for this can be seen in Fig. 9 (bottom) where the $mser_{rel}$ has a larger slope for small and medium cr_t while the $mser_{abs}$ is more flat in this region and is in general more noisy as well.

Furthermore, it can be seen that for the MSSIM metric ($mser_{rel}$) in case of single compression the SRCC$_A$ is smaller compared to the other metrics, which is due to the outlier of the MSSIM at $cr_t = 15$ in Fig. 3. However, except for this outlier, the MSSIM outperforms the other four quality metrics

[4]If absolutely necessary because of ground-truth availability, of course uncompressed data is preferred

	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
PSNR	-0.995	-0.995	-1.0	-1.0	-1.0
MSSIM	-0.912	-0.995	-1.0	-1.0	-1.0
NQM	-0.995	-0.995	-1.0	-1.0	-1.0
VSNR	-0.995	-0.995	-1.0	-1.0	-1.0
RFSIM	-0.995	-0.995	-1.0	-1.0	-0.995

	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
PSNR	-0.857	-0.863	-0.929	-0.868	-0.841
MSSIM	-0.923	-0.863	-0.929	-0.868	-0.841
NQM	-0.857	-0.863	-0.929	-0.868	-0.841
VSNR	-0.857	-0.863	-0.929	-0.868	-0.841
RFSIM	-0.857	-0.863	-0.929	-0.868	-0.846

Table 1: SRCC$_A$ Spearman Rank Correlation Coefficient between quality metrics and $mser_{rel}$ (above) as well as $mser_{abs}$ (below) for homogeneous recompression (*jpg*) and the CAHT segmentation.

	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
PSNR	-0.518	-0.527	-0.541	-0.554	-0.560
MSSIM	-0.519	-0.528	-0.541	-0.554	-0.560
NQM	-0.518	-0.527	-0.541	-0.554	-0.561
VSNR	-0.519	-0.527	-0.541	-0.554	-0.560
RFSIM	-0.519	-0.527	-0.541	-0.554	-0.560

	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
PSNR	-0.054	-0.032	-0.055	-0.035	-0.039
MSSIM	-0.054	-0.032	-0.055	-0.035	-0.038
NQM	-0.053	-0.032	-0.055	-0.035	-0.039
VSNR	-0.054	-0.032	-0.055	-0.035	-0.038
RFSIM	-0.054	-0.032	-0.055	-0.035	-0.038

Table 2: SRCC$_I$ Spearman Rank Correlation Coefficient between quality metrics and $mser_{rel}$ (above) as well as $mser_{abs}$ (below) for homogeneous recompression (*jpg*) and the CAHT segmentation.

and in general represents the segmentation error rates best. In case of homogeneous recompression MSSIM agrees clearest with the trend already seen in section 4.3 (observation (1b)) that for $cr_t > 50$ the segmentation of the single-compressed data shows a lower error and therefore a higher quality than the recompressed data.

Table 2 shows the corresponding SRCC$_I$ results (homogeneous recompression and CAHT segmentation). The 95% confidence intervals are obtained $\pm$ 0.012 – 0.013 for $mser_{rel}$ and 0.017 – 0.019 for $mser_{abs}$. When replacing the mean in the table by the median, the values are increased by $\approx$ 0.07, but the overall trend is identical.

We clearly notice a very different behaviour as compared to the SRCC$_A$ results. While for $mser_{rel}$ values indicate at least linear relation with medium extent, no correlation at all is detectable for $mser_{abs}$. While lower correlation was expected for SRCC$_I$ as compared to SRCC$_A$, the vanishing of correlation for $mser_{abs}$ is surprising. When looking into the single SRCC values in detail it turns out that while $mser_{rel}$ exhibits many values equal to zero[5] which lead to reasonable SRCC$_I$ when correlated with high quality values, $mser_{abs}$ has almost all values > 0 randomly fluctuating[6], which leads to very poor correlation in medium and high quality areas.

[5]A $mser_{rel} = 0$ is common because this only means that the very same segmentation results are achieved in the uncompressed and the compressed image.

[6]This is because already the segmentation results of the uncompressed im-

	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
PSNR	-0.697	-0.785	-0.412	-0.593	-0.275
MSSIM	-0.697	-0.934	-0.385	-0.549	-0.368
NQM	-0.697	-0.934	-0.385	-0.538	-0.368
VSNR	-0.697	-0.934	-0.385	-0.549	-0.368
RFSIM	-0.697	-0.934	-0.385	-0.538	-0.368
	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
NIQE	0.698	0.934	0.385	0.538	0.368
BRISQUE	0.698	0.934	0.385	0.538	0.368
NRPQA	-0.698	-0.934	-0.385	-0.538	-0.368
ANISOTROPY	0.654	0.890	0.093	0.488	0.291

Table 3: SRCC$_A$ between full-reference (above) and non-reference (below) quality metrics and $mser_{abs}$ for heterogeneous recompression (*jxr*) and the WAHET segmentation.

In case of heterogeneous recompression for *j2k*, however, no quality metric describes the behaviour of the mean segmentation errors well. Fig. 10 shows here the best example and clearly differs from the MSSIM graph in Fig. 3. For the *jxr* compression, MSSIM is again the best choice in representing the segmentation error rates since it shows the most linear behaviour and therefore correlates better with the $mser_{rel}$ in Fig. 10 than the other metrics. So in general, MSSIM outperforms the other general purpose quality metrics and agrees with the global trend of the segmentation error rates. However, this metric might be not a sufficient choice in all scenarios, especially if a metric should describe more detailed behaviour than just the global trend of iris segmentation as it could be observed in the *j2k* heterogeneous recompression.

Table 3 shows the SRCC$_A$ correlation results for heterogeneous recompression (*jxr*) and the WAHET segmentation algorithm, comparing full-reference metrics and non-reference metrics. This setting is the one exhibiting lowest SRCC$_A$ values overall. The inconsistent results of (*jxr*) as compared to the other two compression schemes are clearly shown for both types of metrics (contrasting to these results, the SRCC$_A$ values across all settings for *j2k* are always > 0.87 and also for *jpg* correlations are much more consistent compared to *jxr*). It is also clearly visible, that both metric types exhibit similar strengths and weaknesses wrt. low correlations, where non-reference metrics are slightly worse.

As already observed before, SRCC$_I$ values are rather poor for $mser_{abs}$ which is also the case for heterogeneous recompression (*jxr*) and the WAHET segmentation when compared to Table 3. Mean SRCC$_I$ values are found in the interval $[-0.129, 0.120]$ for all metrics (however, still clearly better as compared to $mser_{abs}$ SRCC$_I$ values for *jpg* and CAHT as shown in Table 2), where the 95% confidence interval is around $\pm$ 0.014 – 0.018 the mean values. NIQE and ANISOTROPY exhibit very low mean values even down to 0.027 (i.e. ANISOTROPY for pre-compression quality $q_p = 70$). Considering the median value instead of the means improves the interval to $[-0.175, 0.169]$ with lowest values also seen for NIQE and ANISOTROPY.

ages have a significant $mser_{abs}$. As the segmentation result changes for different qualities no clear trend for the better or worse matching of the ground truth is observable.

	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
jpg	0.703	0.835	0.890	0.786	0.863
j2k	0.978	0.962	0.978	0.923	0.978
jxr	0.742	0.956	0.423	0.544	0.412
	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
jpg	0.863	0.802	0.928	0.868	0.841
j2k	0.984	1.0	0.995	1.0	1.0
jxr	0.978	0.973	0.967	0.918	0.978

Table 4: SRCC$_A$ between $mser_{rel}$ and $mser_{abs}$ for all three methods and both WAHET segmentation (above) and CAHT segmentation (below).

	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
jpg	0.338	0.361	0.379	0.377	0.365
j2k	0.376	0.401	0.384	0.376	0.369
jxr	0.260	0.285	0.279	0.256	0.261
	single	*rec.70*	*rec.75*	*rec.80*	*rec.100*
jpg	0.080	0.051	0.086	0.057	0.058
j2k	0.282	0.295	0.312	0.294	0.282
jxr	0.169	0.169	0.173	0.160	0.175

Table 5: SRCC$_I$ between $mser_{rel}$ and $mser_{abs}$ for all three compression methods and both WAHET segmentation (above) and CAHT segmentation (below).

The SRCC$_A$ between the relative and absolute segmentation error rates confirms that both metrics follow the same trend as it can be seen in Table 4. In case of the WAHET segmentation *j2k* outperforms the other two compression methods. The smaller SRCC$_A$ values for *jpg* are mainly due to the $mser_{abs}$ at smaller compression ratios where the segmentation error rate is higher for $cr_t = 15$ than for $cr_t = 20$. Also the small ripple at $cr_t = 30$ has an impact on the correlation here. The reason for the intermediate SRCC$_A$ results in case of *jxr* are mainly due to noise.

Table 5 shows the corresponding results for SRCC$_I$. The 95% confidence intervals are obtained $\pm$ 0.021 – 0.024 for WAHET and $\pm$ 0.026 – 0.028 for CAHT. When replacing the mean in the table by the median, the values are increased by ≈ 0.1 – 0.18, but the overall trend is identical.

When compared to SRCC$_A$ values, we notice very different behaviour. SRCC$_I$ suggests better correlation for WAHET compared to CAHT, where for both segmentation algorithms *j2k* leads to highest correlation (though on a much lower level as suggested by SRCC$_A$ results). *jxr* is worse compared to *jpg* for WAHET, while *jpg* is clearly worst in terms of SRCC$_I$ for CAHT (which confirms the poor correlation of $mser_{abs}$ to the quality metrics as shown in Table 2). Results suggest that CAHT in general tends to produce more fluctuating results for $mser_{abs}$ which impact SRCC$_I$ but are averaged out for SRCC$_A$.

Overall, we are able to demonstrate reasonable correlation between image quality metrics and image segmentation errors as long as averaged metrics values are correlated to averaged segmentation error values. In case single image correlations are considered and averaged subsequently, correlation drops significantly and specific combinations (like *jpg* compression and CAHT segmentation) lead to highly fluctuating $mser_{abs}$ error values causing correlations to vanish entirely. Thus, the quality of a single image cannot be used as a reliable predictor

for the expected segmentation error, especially in problematic cases like the ones described above.

6. Recompression Effects on Recognition Accuracy

Iris segmentation is only one component of the entire iris recognition pipeline. Thus, for assessing recompression effects on iris recognition systems, experiments need to be conducted considering the entire pipeline. In this context, it is crucial to use different iris recognition schemes since it can be expected that different feature extraction strategies will react differently when being confronted with compression artifacts and reduced image quality in general. Thus, in addition to the two segmentation schemes used before, we employ custom implementations of four feature extraction techniques (for a description of our implementation of preprocessing, feature extraction, and matching see [2]). All implementations are available in USIT (University of Salzburg Iris-Toolkit at `http://www.wavelab.at/sources`).

The first scheme has been developed by Ko et al. and extracts spatial domain features, while the second approach has been designed by Monro et al. and relies on DCT-derived features computed from rotated texture patches. The third scheme has been published by Ma et al. using a 1D dyadic wavelet transform maxima representation for small averaged stripes of the iris texture while the fourth technique is a re-implementation of the popular 1D log-Gabor MATLAB-code of Libor Masek.

For evaluation and computing the receiver operating characteristic (ROC, which is used to rate and compare different configurations), the protocol as suggested for the fingerprint verification contests (FVC [37]) is used: While all genuine scores are computed, the impostor score set is generated using the first sample of each eye only to result in a more balanced number of genuine and imposter score matches. See Table 6 for equal error rates (EERs) on uncompressed data (serving as "ground truth").

	Masek	Ma	Monro	Ko
CAHT	1.86	1.75	1.74	3.00
WAHET	6.65	6.88	6.74	9.40

Table 6: EER (%) of various settings on uncompressed data.

Depending on the data iris texture feature extraction is being applied to, two different compression scenarios can be distinguished: The compressed vs. compressed case, denoted as CCC, where both templates involved in matching, the one generated from the sample data and the one from the database, are derived from images compressed to the same bitrate. The compressed vs. uncompressed case, denoted as CUC, where the template generated from the compressed sample is matched against the database containing templates derived from uncompressed iris images.

In the following plots, the x-axis depicts compression rate while the EER is shown on the y-axis. In each plot, the EERs for single compression and four recompression settings ($q_p = 70, 75, 80, 100$) are plotted against increasing compression rate. Fig. 11 compares results of the two compression scenarios CUC and CCC.

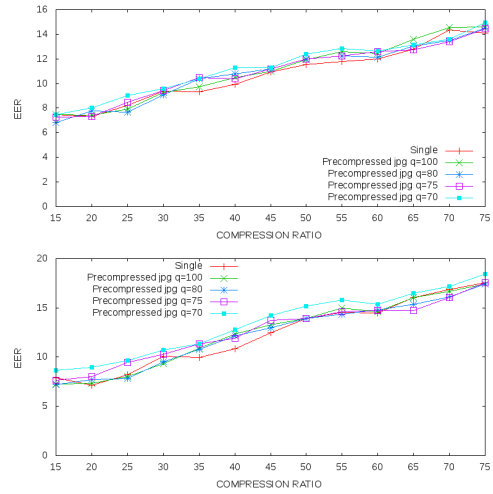


Figure 11: Comparing compression scenarios for *j2k*, WAHET segmentation, and Masek feature extraction. Top: CUC, Bottom: CCC.

The first obvious thing to note is that EERs are consistently lower for the CUC scenario and of course, EERs are already increased at $c_r = 15$ as compared to the uncompressed case (compare Table 6). But also with respect to recompression effects we note a difference. For $c_r \leq 60$, pre-compression with $q_p = 70$ and $q_p = 75$ leads to clearly worse EERs as compared to single compression for CCC. A similar trend can be observed for CUC, but it is less clear. This corresponds to very general overall observations: First, we observe more distinct differences between single compression and recompression in the CCC scenario. Second, even more general, for settings resulting in worse EERs, we observe more distinct differences between single compression and recompression as well.

Fig. 12 investigates the effect of different segmentation schemes while keeping the other parameters fixed (i.e. Ko feature extraction, CUC scenario, *jxr*).

Again, for both segmentation schemes, heterogeneous recompression with $q_p = 70$ and $q_p = 75$ tends to deliver worse EERs as compared to single compression, however, more clearly so for WAHET. This corresponds to the above statement about more distinct differences for lower EERs in general. In Fig. 13 we consider homogeneous *jpg* recompression using two different feature extraction schemes. While the homogeneous recompression case exhibits the lowest differences between single compression and recompression, this example shows entirely contrasting behaviour at $c_r = 30$: While for Masek feature extraction recompression with $q_p = 70$ is actually better than single compression, for Monro feature extraction the opposite is true and $q_p = 75$ is even worse.

In Fig. 14 we provide additional experimental evidence that in some settings, heterogeneous recompression with $q_p = 70$ and $q_p = 75$ delivers worse EERs as compared to single com-

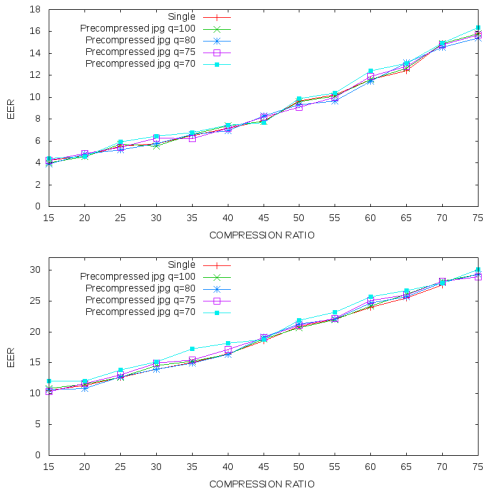


Figure 12: Comparing segmentations for *jxr*, Ko feature extraction in CUC scenario. Top: CAHT, Bottom: WAHET.

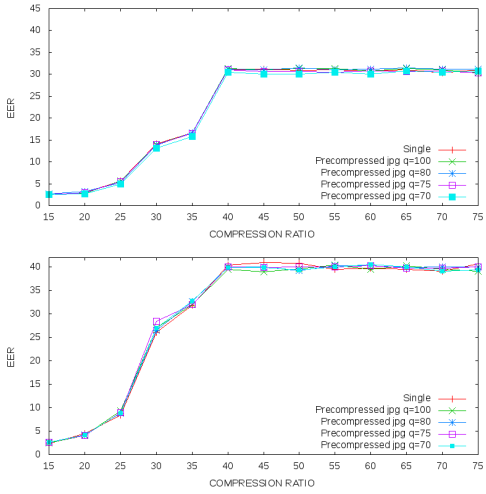


Figure 13: Settings with contrasting effects for *jpg*, CAHT segmentation in CCC scenario. Top: Masek, Bottom: Monro.

pression.

From the results we observe the following:

1. Results are not as clear as for assessing segmentation errors only. In many compression, segmentation, and feature extraction combinations the differences between single compression and recompression seem to be random and even entirely differ among "adjacent" compression rates.
2. While in the assessment of segmentation error dependency on various (re)compression settings homogeneous *jpg* recompression gives the clearest (negative) results for high compression rates, this type of recompression hardly influences overall recognition performance. It seems that the compressions' impact on iris texture quality dominates

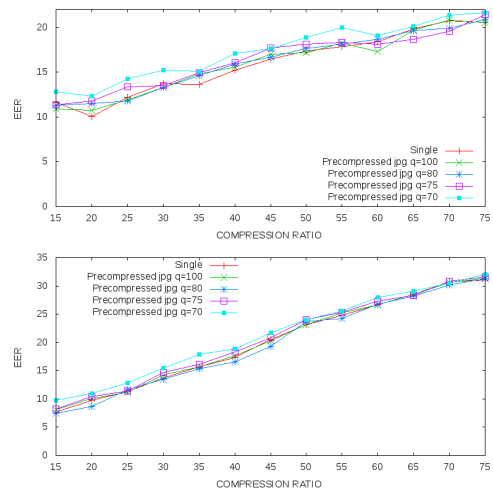


Figure 14: Further examples for better values of single compression compared to recompression in CCC scenario and WAHET segmentation. Top: *j2k* and Ko, Bottom: *jxr* and Ma.

eventual negative effects in segmentation. Heterogeneous recompression leads to more impact on recognition accuracy.

3. There is a trend that settings leading to lower overall EERs exhibit more distinct differences between single compression and recompression (CUC vs. CCC; CAHT vs. WAHET; Ma, Masek, and Monro vs. Ko). However, this is no strict rule and overall it is impossible to predict for which settings relevant differences are found.
4. The largest differences observed are in the range of increasing EER by 25% when comparing recompression and single compression. However, it has to be clearly stated that such large differences are rarely seen. In most cases, differences are small and seem to be randomly fluctuating.

However, the EER only represents one specific point on the ROC curve corresponding to a specific user convenience / system security relation. In order to consider a second corresponding relation, we investigate the situation in terms of the smallest FRR, for which the FAR=0%. This corresponds to a high security setting without any false accepts. Thus, in the following plots, the x-axis depicts compression ratio while the FRR (%) at FAR=0 is shown on the y-axis. In each plot, the FRRs for single compression and four recompression settings ($q_p = 70, 75, 80, 100$) are plotted against increasing compression rate. Fig. 15 compares results of the two compression scenarios CUC and CCC.

Results confirm the observation for ERRs that in the CUC scenario, results are clearly better. We again observe larger differences among the different recompression settings in the CCC scenario case, however, these differences do not seem to be systematic but are fluctuating randomly. Fig. 16 shows the same comparison replacing *j2k* with *jxr* heterogeneous recompression. Again, the CCC scenario is clearly worse up to a compression ratio of 55. The differences between the two scenarios

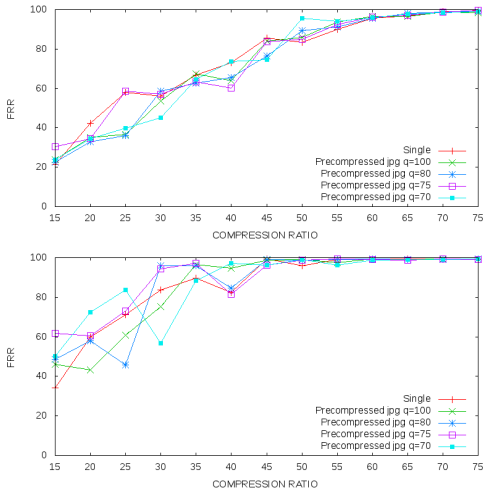


Figure 15: Comparing compression scenarios for *j2k*, CAHT segmentation, and Monro feature extraction (FRR @ FAR=0). Top: CUC, Bottom: CCC.

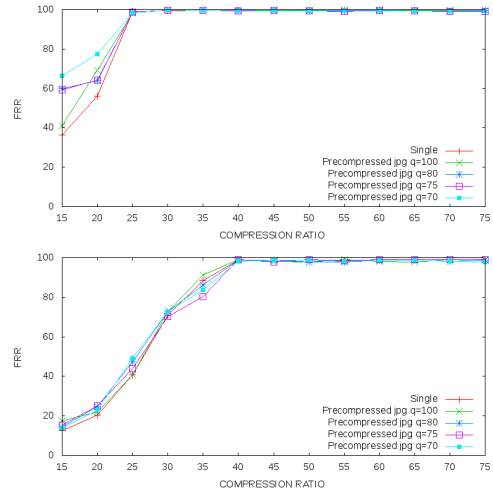


Figure 17: Further examples for better values of single compression compared to recompression in CCC scenario and *jpg* compression (FRR @ FAR=0). Top: WAHET and Monro, Bottom: CAHT and Ko.

are observed on a rather regular basis throughout the results, however the extent of difference is highly varying.

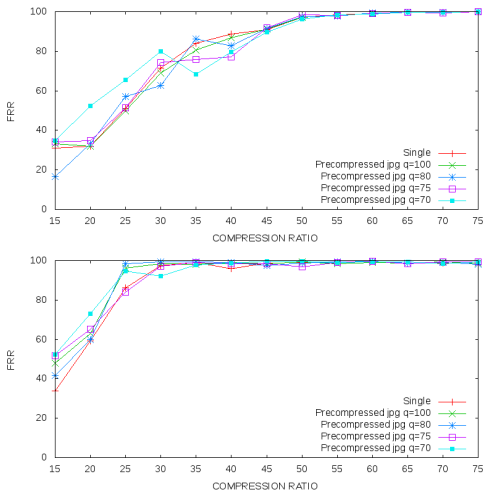


Figure 16: Comparing compression scenarios for *jxr*, CAHT segmentation, and Monro feature extraction (FRR @ FAR=0). Top: CUC, Bottom: CCC.

Fig. 16 illustrates one of the rare cases where single compression is clearly superior to recompression in terms of FRR – for the CUC scenario only at compression ratios 20 and 25 (and very clear compared to recompression with $q_p = 70$ only), but for CCC this is also true for compression ratio 15 and several recompression settings. The same effect (i.e. superiority for single compression at some compression ratios (ratios 15 and 20 for the top plot and 15 to 30 for the lower plot) for several recompression settings) is seen in Fig. 17.

FRR @ FAR=0 results basically confirm results seen when considering EER. While the differences between the two compression scenarios CUC and CCC are clear and systematic, we find cases of heterogeneous recompression where we observe differences between single compression and recompression but these differences occur almost randomly and do not follow a strict and predictable pattern. Having seen all these differences the question arises naturally if the observed differences are statistically significant. While it is almost impossible to come up with corresponding results for all possible configurations, we follow the strategy proposed in [38] and provide bounds for significant differences. Fig. 18 displays the required Δ according to [38] in terms of EER (left) and FRR (right – note that the methodology in [38] may be easily extended to cover FRR as well but leads to much larger required differences due to the missing impostor matches) to result in a significant difference in case the larger EER_M or FRR_M in a comparison is attained as given on the x-axis.

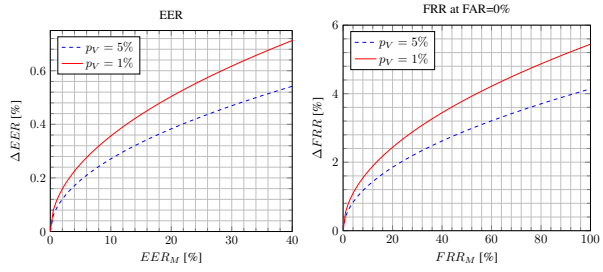


Figure 18: Visualisation of the required Δ in terms of ERR (left) and FRR (right) to result in a significant difference.

The values shown in Fig. 18 imply e.g. for Table 6 (comparing segmentation and feature extraction techniques on uncompressed data), that WAHET results are significantly different from CAHT ones for identical feature extraction, and that Ko feature extraction is significantly different form the other

feature extraction schemes, while Masek, Ma, and Monro feature extraction schemes are not significantly different from each other when used with identical segmentation. With respect to the observed differences in terms of compression settings CUC and CCC (Fig. 11 for EER and Figs 15 as well as 16 for FRR) most of the examples turn out to describe significant ones. The same is true for the differences among different recompression settings in all figures shown – for most differences clearly visible in the figures, a significance even at level $p_V = 1\%$ can be stated, for EER as well as for FRR @ FAR=0 results.

7. Conclusion

Often researchers use JPEG pre-compressed data for iris biometrics compression performance testing in experimental experiments, mostly because public data-set are often only available in this format. We investigated if the outcome of such experiments can be considered reliable by comparing segmentation error, image quality metrics, and recognition performance of single-compressed and recompressed data. In the segmentation error rate, no tendency is observable when comparing single-compression and heterogeneous recompressed data. However, using homogeneous recompressed data, i.e. compressing JPEG pre-compressed data with JPEG again, a different behaviour is observed for high compression ratios compared to single-compressed data sets. Thus results of studies using JPEG compression on JPEG pre-compressed data cannot be considered entirely reliable when focus is set on segmentation. We further showed for small compression ratios, a ground truth is indeed necessary for accurate segmentation error rating. We also propose a method to overcome such problems in section 4.3. Interestingly, there is no strictly linear relation between image quality and segmentation error rate. Quality metrics tend to omit detailed behaviour of the segmentation error wrt. compression ratios. Nevertheless, quality metrics and segmentation error follow the same trends and quality metrics can therefore be used to estimate an iris segmentation algorithms behaviour as long as averaged metrics values and segmentation errors of an entire dataset are being used while single image results can be misleading. When the entire iris recognition pipeline is considered, the picture changes somewhat. In homogeneous JPEG recompression we hardly find significant impact on recognition accuracy when comparing single-compressed and recompressed data. In heterogeneous recompression applying JPEG XR and JPEG2000 to JPEG pre-compressed data, we indeed find cases where recognition accuracy is lower for recompressed data. However, for a given setting it is impossible to predict these differences. This implies that from the overall recognition accuracy viewpoint, studies using any other compression scheme apart from JPEG on JPEG pre-compressed data cannot be considered entirely reliable and need to be verified using uncompressed data in experimentation.

References

[1] M. Burge, K. Bowyer (Eds.), Handbook of Iris Recognition, Springer Verlag, 2013.

[2] C. Rathgeb, A. Uhl, P. Wild, Iris Recognition: From Segmentation to Template Security, Vol. 59 of Advances in Information Security, Springer Verlag, 2013.

[3] S. Rakshit, D. Monro, An evaluation of image sampling and compression for human iris recognition, IEEE Transactions on Information Forensics and Security 2 (3) (2007) 605–612.

[4] R. W. Ives, R. P. Broussard, L. R. Kennell, D. L. Soldan, Effects of image compression on iris recognition system performance, Journal of Electronic Imaging 17 (2008) 011015, doi:10.1117/1.2891313.

[5] S. Matschitsch, M. Tschinder, A. Uhl, Comparison of compression algorithms' impact on iris recognition accuracy, in: S.-W. Lee, S. Li (Eds.), Proceedings of the 2nd International Conference on Biometrics 2007 (ICB'07), Vol. 4642 of LNCS, Springer Verlag, 2007, pp. 232–241.

[6] J. Hämmerle-Uhl, C. Prähauser, T. Starzacher, A. Uhl, Improving compressed iris recognition accuracy using JPEG2000 RoI coding, in: M. Tistarelli, M. Nixon (Eds.), Proceedings of the 3rd International Conference on Biometrics 2009 (ICB'09), Vol. 5558 of LNCS, Springer Verlag, 2009, pp. 1102–1111.

[7] M. Konrad, H. Stögner, A. Uhl, Custom design of JPEG quantization tables for compressing iris polar images to improve recognition accuracy, in: M. Tistarelli, M. Nixon (Eds.), Proceedings of the 3rd International Conference on Biometrics 2009 (ICB'09), Vol. 5558 of LNCS, Springer Verlag, 2009, pp. 1091–1101.

[8] J. Daugman, C. Downing, Effect of severe image compression on iris recognition performance, IEEE Transactions on Information Forensics and Security 3.

[9] R. W. Ives, D. Bishop, Y. Du, C. Belcher, Iris recognition: The consequences of image compression, EURASIP Journal of Advances in Signal Processing 2010 (2010) Article ID 680845, doi:10.1155/2010/680845.

[10] P. Philips, K. Bowyer, P. Flynn, Comments on the CASIA version 1.0 iris data set, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (10) (2007) 1869–1870.

[11] K. Horvath, H. Stögner, A. Uhl, Effects of JPEG XR compression settings on iris recognition systems, in: P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano, W. Kropatsch (Eds.), Proceedings of the 14th International Conference on Computer Analysis of Images and Patterns (CAIP 2011), Vol. 6855 of LNCS, Springer Verlag, 2011, pp. 73–80.

[12] A. Tuba, J. Arezina, R. Jovanovic, Software framework for testing JPEG quantization tables for iris recognition, in: Applied Mathematics in Electrical and Computer Engineering, 2012, pp. 417–422.

[13] B. Pardamean, I. Christiani, Comparatative evaluation of iris recognition performance, International Journal of Mathematical Models and Methods in Applied Sciences 6 (2) (2012) 332–339.

[14] M. Carneiro, A. Veiga, E. Flores, G. Carrijo, Solutions for iris segmentation, in: Z. Riaz (Ed.), Biometric Systems, Design and Applications, InTech, 2011, pp. 111–128.

[15] B. Klare, M. Burge, Assessment of H.264 video compression on automated face recognition performance in surveillance and mobile video scenarios, in: B. Kumar, S. Prabhakar, A. Ross (Eds.), Biometric Technology for Human Identification VII, Vol. 7667 of SPIE Proceedings, 2010, p. 76670X.

[16] P. Kumar, M. Prasad, K. Padmaja, Face recognition of database of compressed images using local binary patterns, International Journal of Computer Applications 74 (16) (2013) 10–17.

[17] W. Pennebaker, J. Mitchell, JPEG – Still image compression standard, Van Nostrand Reinhold, New York, 1993.

[18] S. Chan, Recompression of still images, Tech. Rep. 21068, Computing Laboratory, University of Kent, UK (1992).

[19] B. Kumar, G. Karpagam, An empirical analysis of requantization errors for recompressed jpeg images, International Journal of Engineering Science and Technology 3 (12) (2011) 8519–8527.

[20] H. Bauschke, C. Hamilton, M. Macklem, J. S. McMichael, N. Steward, Recompression of jpeg images by requantization, IEEE Transactions on Image Processing 12 (7) (2003) 843–849.

[21] H. Sencar, N. Memon, Digital Image Forensics: There is more to a picture than meets the eye, Springer Verlag, 2012.

[22] C. Rathgeb, A. Uhl, P. Wild, Effects of severe image compression on iris segmentation performance, in: Proceedings of the IEEE/IAPR International Joint Conference on Biometrics (IJCB'14), 2014. doi:10.1109/ICB.2013.6613010.

[23] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, A. Uhl, A ground

truth for iris segmentation, in: Proceedings of the 22th International Conference on Pattern Recognition (ICPR'14), Stockholm, Sweden, 2014, p. 6pp.

[24] T. Bergmüller, E. Christopoulos, M. Schnöll, A. Uhl, Recompression effects in iris segmentation, in: Proceedings of the 8th IAPR/IEEE International Conference on Biometrics (ICB'15), Phuket, Thailand, 2015, pp. 1–8.

[25] K. W. Bowyer, P. J. Flynn, The nd-iris-0405 iris image dataset, Tech. rep., University of Notre Dame.

[26] S. Bharadwaj, M. Vatsa, R. Singh, Biometric quality: a review of fingerprint, iris, and face, EURASIP Journal on Image and Video Processing 2014:34.

[27] J. Hämmerle-Uhl, M. Pober, A. Uhl, General purpose bivariate quality-metrics for fingerprint-image assessment revisited, in: Proceedings of the IEEE International Conference on Image Processing (ICIP'14), Paris, France, 2014.

[28] Z. Wang, E. P. Simoncelli, A. C. Bovik, Multi-scale structural similarity for image quality assessment, IEEE Asiloma Conference on Signal, Systems and Computers.

[29] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, A. C. Bovik, Image quality assessment on a degradation model, IEEE Transactions on Image Processing, vol. 9, no. 4, pp. 636-649.

[30] X. M. L. Zhang, L. Zhang, RFSIM: A feature based image quality asesment metric using Riesz transforms, Proc. 17th IEEE Int. Conf. on Image Procesing ICIP (2010) p. 321–324Hong Kong, China.

[31] D. M. Chandler, S. Hemami, Vsnr: A wavelet-based visual signal-to-noise ratio for natural images, IEEE Transactions on Image Processing vol. 16, no. 9 (2007) pp. 2284–2298.

[32] S. Gabarda, G. Cristobal, Blind image quality assessment through anisotropy, Journal of the Optical Society of America. A, Optics, image science, and vision 24 (12) (2007) B42–51.

[33] Z. Wang, H. Sheikh, A. Bovik, No-reference perceptual quality assessment of JPEG compressed images, in: Proceedings of the 2002 IEEE International Conference on Image Processing (ICIP'2002), 2002, pp. I–477 – I–480.

[34] A. Mittal, A. Moorthy, A. Bovik, No-reference image quality assessment in the spatial domain, IEEE Transactions on Image Processing 21 (12) (2012) 4695 – 4709.

[35] A. Mittal, R. Soundararajan, A. C. Bovik, Making a completely blind image quality analyzer, IEEE Signal Processing Letters 22 (3) (2013) 209–212.

[36] C. Rathgeb, A. Uhl, P. Wild, Effects of severe image compression on iris segmentation performance (best poster award), in: Proceedings of the IAPR/IEEE International Joint Conference on Biometrics (IJCB'14), 2014.

[37] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, A. K. Jain, FVC2004: Third Fingerprint Verification Competition, in: ICBA, Vol. 3072 of LNCS, Springer Verlag, 2004, pp. 1–7.

[38] H. Hofbauer, A. Uhl, Calculating a boundary for the significance from the equal-error rate, in: Proceedings of the 9th IAPR/IEEE International Conference on Biometrics (ICB'16), 2016, pp. 1–4.