Tamper-Resistant Biometric IDs

Darko Kirovski, Nebojša Jojić, and Gavin Jancke
Microsoft Research, One Microsoft Way, Redmond, WA, 98052, USA
{darkok,jojic,gavinj}@microsoft.com

Abstract

We present FaceCerts, a simple, inexpensive, and cryptographically secure identity certification system. A FaceCert is a printout of person's portrait photo, an arbitrary textual message, and a 2-D color barcode which encodes an RSA signature of the message hash and the compressed representation of the face encompassed by the photo. The signature is created using the private key of the party issuing the ID. ID verification is performed by a simple off-line scanning device that contains the public key of the issuer. The system does not require smart cards; it can be expanded to encompass other biometric features, and more interestingly, the ID does not need to be printed by a trusted or high-end printer, it can be printed anywhere, anytime, and potentially by anyone. The ID verifier uses a single scan process which does not require the use of displays. We detail system's components and present a preliminary performance evaluation using an in-field experiment.

1 Introduction

A typical identity certification such as a driver's licence, passport, or visa, consists of a personal portrait photo, an arbitrary message, and one or more features whose purpose is to guarantee authenticity. Commonly, authenticity is assured using sophisticated printing procedures that are difficult to replicate: holograms, watermarks, micro-printing and threading, special print paper, and chemical coating [1]. However, the wide availability of such technology has rendered forging most personal ID documents a relatively simple task with results often perceptually comparable to the originals. Authentication of imprinted features via electronic devices is complex and most importantly, expensive [1].

In all-digital environments such as smart cards or lasercards [2], authenticating the source of a personal ID is an easy task using off-the-shelf public cryptography [3] and one-way authentication protocols [4]. Typically, the stored photograph as well as other biometric features are concatenated to the textual message and hashed. The resulting hash is then signed using the private key of the issuer. In-field authentication is performed using the public key of the issuer by a verification device (e.g., smart card reader), which also must display the signed data. While the security of such systems can be made to follow even the strictest security standards, the cost of supporting systems makes them undesirable for widespread identity certificate applications such as national ID cards, driver's licences, or passports. A simple smart card costs about \$5-\$35, while a lasercard reader costs about \$2400 [5].

In this paper, we combine best of both worlds into a new technology we call FaceCerts and show how sophisticated specialized compression algorithms can allow the use of paper as an inexpensive hybrid analog/digital domain on which both the human readable information, i.e., text and photo, and the secure digital information can be stored in a way that allows a single-scan verification.

1.1 FaceCerts

Instead of relying on the sophistication of the printing process to impose difficult forging, FaceCerts rely on public-key cryptography for provable security, while deploying a standard-quality low-cost color printing process which keeps the cost of printing a FaceCert two orders of magnitude lower than that of a smart card or a lasercard. Issuing and verification of FaceCerts is illustrated[1] in Figure 1.

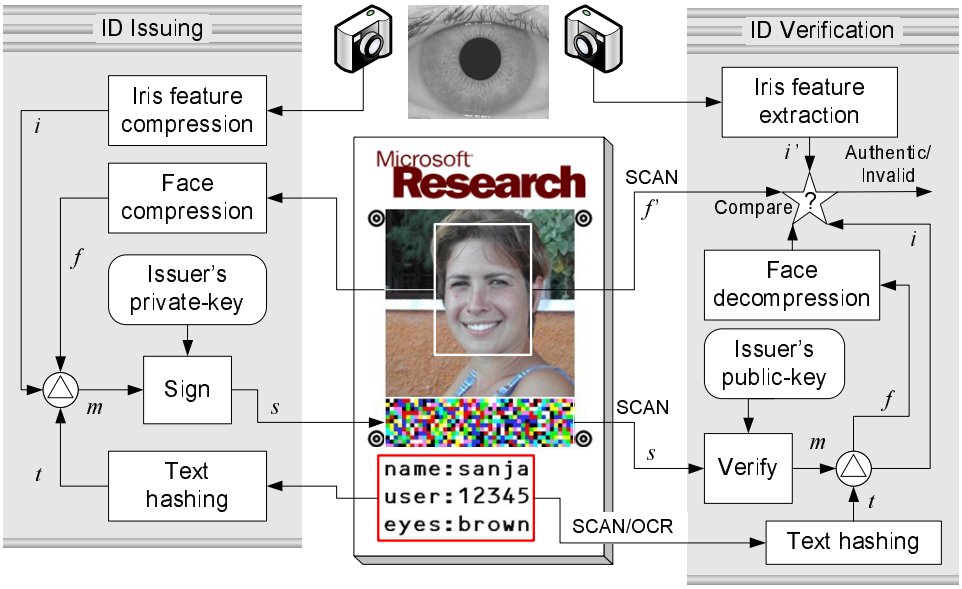

Figure 1: Functional block diagram of the actions taken at the issuer and verifier of FaceCert IDs.

The information certified on a FaceCert is both biometric and textual. The biometric data can encompass facial, iris, and/or other features. The digital photo that represents a portrait of the FaceCert holder, is the only biometric feature printed in plain-text on the ID. The textual data can be of arbitrary length and is also printed on the ID. The ID is certified in the following way. First, the textual data is hashed using a cryptographically secure hashing algorithm such as SHA1 [7]. The resulting 160-bit hash is denoted as t. Next, the facial features on the photo are identified and compressed using an algorithm partially described in this manuscript in Section 3. The best-effort output of the face compression step, denoted as f, is constrained to 1-2Kbits. The actual compression rate can be adjusted to meet the desired balance of picture quality vs. barcode size, which is mandated by the application. Picture quality affects system performance for two main reasons: first, to impose low likelihood of a false negative or positive during detection and second, to set the desired level of facial feature detail which an adversary, whose photo has not been taken for the ID, must resemble in order to use the authentic FaceCert.

Other biometric information such as iris or fingerprint patterns can be also certified and verified using a FaceCert ID as illustrated in Figure 1. For example, details of a FaceCert system that encompasses person authentication via iris patterns is described in [8]. We omit the details of the feature extraction and compression process for iris patterns in this manuscript; instead, we recognize a string of bits, i, as an output of this process. The iris digest can be

[1]The iris scan in this figure has been taken from the CASIA iris database. Portions of the research in this paper use the CASIA iris image database collected by Institute of Automation, Chinese Academy of Sciences [6].

typically compressed down to 700-1.5Kbits [8] depending upon the desired balance of error rates vs. barcode size.

Messages $f||i$ [2] and t are merged into a message $m = (f||i)\triangle\ t$ using a reversible non-commutative operator $\triangle$ such that $(\exists\triangle^{-1})f||i = m\triangle^{-1}t$. Note that $f||i$ are not hashed because their plain-text values must be retrieved during FaceCert verification. Also it is safe to assume that message t can be recovered error-free from a printed FaceCert because of deployed high-performance optical character recognition engines [9], [10]. Shortly, we review this operator in more detail.

In the next step, message m is signed with the private key of the FaceCert issuer. We use an RSA private key of $|m| + 1$ bits to sign/decrypt m. Considering typical lengths of f and i, we bound the length of m within $1300 < |m| < 3400$ bits. The resulting signature s is encoded using Reed-Solomon error correction codes [11] and printed as a barcode onto the FaceCert.

Two aspects of printing and scanning are important: *degradation of printed color* and *scanning reliability*. Independent studies have shown that state-of-the-art inks have an estimated life of 65 years on a cotton paper in average indoor display without noticeable fading and several years of corresponding outdoor lifetime [12]. The second requirement has been already addressed in modern barcode standards such as PDF417 [13].

A FaceCert verifier initially scans all three printed components: the photo, the text, and the barcode. The barcode is decoded into the originally printed signature s. The scanned textual data is also converted into a text-string using reliable optical character recognition. For successful verification of a FaceCert, the text and the barcode need to be read without errors. Next, after verifying/encrypting the signature with the corresponding public RSA key of the issuer [3], the verifier obtains the signed message m. After the verifier hashes the text to obtain t, it computes $f||i = m\triangle^{-1}t$. Then, the verifier decompresses f into a subimage of the original photo that contains the facial features. Finally, the verifier quantifies the level of similarity between the decompressed and scanned face. If the two images are similar within the maximum tolerable compression-print-scan noise, only then the FaceCert is declared as authentic.

In case additional biometric features such as iris patterns are required for person authentication, the FaceCert verifier captures a photo of person's iris, extracts its features, and compares them to the features decompressed from i. If the feature comparison yields positive identification, the FaceCert is declared as authentic. Since the FaceCert verifier does not query a trusted database with iris digests (i.e., does not perform the traditionally error-prone iris recognition procedure), the detection threshold in the FaceCert system can be set to adjust for much lower false positive error rates than "classic" iris recognition systems [8]. In general, two verification procedures are recognized in the FaceCert system: (a) low-cost, where only facial features are verified, and (b) high-security, where both facial and iris (possibly, fingerprint and retina) patterns are verified in order to decide upon FaceCert's authenticity.

Finally, we revisit the selection of the operator $\triangle$. Its purpose is to prevent adaptive existential forgery on the signing primitive, e.g., RSA, where the adversary creates a valid signature with no control over the message [14], [15], [16]. This problem is well known to the cryptography community and has been addressed in several protocols including the probabilistic signature scheme with message recovery (PSS-R) [17], which is based upon optimal asymmetric encryption padding (OAEP) [18]. Although several integrity check mechanisms for RSA signatures can be used with different security properties, the exemplary PSS-R achieves provable security with near-optimal redundancy used in order to achieve a desired level of security. In case $\triangle$ = PSS-R, then message m is created by setting $M = f||i$ and hashing $M||t||r$ to obtain $w = h(M||t||r)$,

[2] Operator || denotes concatenation.

where $h()$ is a hash function and M, w, and r refer to the corresponding variables in Figure 2 in Section 5 of [17]. PSS-R derives $m = b||w||r^*||M^*$ where b is a single bit set to 0 and variables r^* and M^* are created as in Figure 2 in Section 5 of [17]. Signature's integrity check in this case is performed according to the *RecPSSR* procedure presented in Section 5 of [17] with the last step altered to: **if** $h(f||i||t||r) = w$ and $b = 0$ **then return** $f||i$ **else return REJECT**. The signed message m has bit-length $|(f||i)| + 2hLen + 16$, where $hLen$ is the length of the output of the hash function $h()$ in bits (160 bits for SHA1 [7]).

Under the assumption that the cryptographic functions are signing the biometric properties and the associated text in a provably secure manner, the security of FaceCerts stems from the fact that changing a single bit of the textual message or altering the photo beyond the print-scan noise causes a global change in the barcode that appears to be random without the knowledge of the issuer's private key.

In this manuscript, we focus on the two crucial components of the system, a novel face compression algorithm and a statistical metric for computing similarity between an original and a corresponding compressed face in the presence of print-scan noise. The basic requirement for the face compression algorithm in the FaceCert system is to compress an image of a face into only several thousand bits with preserved sharpness of the main facial characteristics. We present a novel face compression technology based on eigenfaces [19] and improved variants of principal component analysis [20], [21]. We show that our technology achieves desired compression rates even when the component analysis is trained on a small database of images.

2 Related Work

The idea of using digital technology and cryptography as key to enabling low-cost photo identification is not new. For example, one centralized card authentication system which relies on displays has been developed and marketed by Kodak [22]. It stores a users photograph on a card in a highly compressed code on the magnetic stripe or smart-card memory. The authentication procedure entails reading the encoded photograph, comparing it against its database entry, and displaying it on a screen for comparison against the cardholder.

System presented by O'Gorman and Rabinovich [23] is the most related to our work as it aims at the same goal - however, it relies on signing *image digests which are tolerant to scanning errors* **instead of** *actual compressed images*. In this manuscript, we show a successful attack on the O'Gorman-Rabinovich system that manipulates an image using a simple procedure so that its digest equals the digest of another distinct facial photograph. By using a compressed version of the facial structure within an image instead of the image digest, in the case of FaceCerts such attacks are reduced to seeking perfect human look-a-likes. Since this is a limitation of the distinctiveness of a human face, the FaceCert system supports additional biometric information such as iris patterns.

2.1 Comparison with Existing Solutions

Biometric Recognition. Another alternative to FaceCerts is biometric recognition. Biometrics has been defined as a process of automatically recognizing a person using distinguishing traits. Several biometric solutions have been proposed via face, speech, fingerprint, handwriting, iris, and retina recognition. Solid survey of these techniques can be found at [24]. Just as FaceCerts, a person identification system that relies on biometric solutions *must* involve a human verifier who must ensure the identification system is not fooled. For instance, an adversary can

show a realistic size photo of the face of an authorized person to the face detector or play a voice recording to a speaker detector.

While some types of biometric identification such as fingerprint detection are reliable, they can be used maliciously to incriminate innocent users [25]. A malicious detector can record a person's fingerprint, create its physical copy, and then, incriminate this person at will. This renders fingerprint detection systems relatively undesirable for most person authentication scenarios. Some biometrics systems are commonly subject to complaints for invasion of privacy [5]; e.g., wide-spread face detection points can disclose at any time one's location to a party who gains control over such a system. Nevertheless, the three most important disadvantages of almost all biometric recognition systems are:

- *reliability*, in particular in face and speaker recognition, does not stay constant as the system scales up, which commonly renders these systems highly prone to false alarms and false positives [26], [27], and
- *centralized decision making* – the verifier needs to be connected to a central trusted server which actually performs the identification, which in a sense implies:
- *high cost* – the equipment performing the verification is costly.

For most applications, such solutions are inconvenient, costly, and most importantly, unreliable.

Smart cards. Smart cards represent an effective solution to person identification. A big advantage of smart cards is all-digital communication with the verification device. A simple scenario is to have a smart card which contains a digital photo, personal biometric and description data, and a signed hash of this information using the private key of the issuer. Verification is performed by hashing the photo and the personal description data and then verifying this hash against the signature using the public key of the issuer. Finally, the verifier *must* display the verified digital photo, so that a human can acknowledge that the person being identified is on the photo. Note that a printed photo on the smart card is ineffective because a malicious party can trivially extract the contents of a valid card, then create another one with the same digital contents however with a different printed photo.

Smart cards just as FaceCerts, cannot be used to store private information (e.g., private keys which are revoked if smart card is lost). It has been demonstrated so far that smart cards cannot be considered a secure storage because it is relatively easy to extract the hidden information even without reverse engineering the smart card [28]. Exemplary attacks that have successfully identified encryption keys (both symmetric and private keys), have been based on analyzing smart card's I/O behavior via differential power analysis [29] or timing analysis [30].

Finally, there are several differences that strongly favor FaceCerts to smart cards.

- A smart card based system must display the photo, whereas FaceCerts only scans the ID with no requirement to display any imagery. Medium-quality displays are significantly more expensive than CCD (charge-coupled device) scanners (up to a factor of 5), which reduces significantly the cost of the verifying infrastructure.
- Personal IDs are frequently lost or damaged. Replacing a FaceCert involves only reprint, whereas replacing a smart card involves purchase of another hardware device in addition to burning this device with the appropriate identification contents – two orders of magnitude differential in replacement cost.

It is important to stress that smart cards should not be understood as competition to FaceCerts; on the contrary, the information printed on a FaceCert can be stored in its digital format on a smart card and verified in an "all-digital manner" without scanning. The main benefit of FaceCerts is that they enable the inexpensive paper ID version.

Watermarks. Another technique for authenticating content is to hide an imperceptible secret information, watermark, in the digital photo. One serious disadvantage of this type of ID authentication is the fact that in most watermarking systems, the secret hidden in the photo must be present in the verifier. Hence, a single broken verifying device renders the entire system broken. A public-key watermarking system has been developed, however, with a different target application [31]. This system requires significantly longer host signals than a single photo to reliably detect the existence of a given secret. Also, such a system would require that the secret used to mark a photo is renewed after issuing several distinct IDs. In summary, using modern watermark-based technologies results in the least robust and secure performance for secure identity certification.

3 FaceCerts - Face Compression

The computer vision community has studied various models of faces in the past. The system we are proposing in this paper does not need to encode the face image to facilitate recognition of the person when observed under various new conditions, such as angle of view and illumination changes, aging, or facial hair changes, but rather in the very same photograph from which the face code has been extracted. Thus, we do not face the difficult issue of over-training that is present in a typical face recognition application. Rather, our needs are simply for a very efficient face image compression.

As faces form a class of images with substantially smaller variability then the class of all natural images, they can be compressed better by using a class-specific compression scheme than using general-purpose compression algorithms, such as JPEG. To develop such a scheme, we need to model the variability of facial images, i.e., the probability distribution $p(\mathbf{g})$, where $\mathbf{g}$ denotes the vector of pixel intensity in a facial image. Then, according to Shannon's coding theorem, the code length for the image g is bounded bellow by $-\log_2 p(\mathbf{g})$ bits. To build this distribution, we focus on 2D subspace models.

The problem of subspace learning can be elegantly defined in terms of a generative model that describes joint generation of the subspace coordinates, or factors, $\mathbf{y}$ and the image $\mathbf{g}$ by linearly combining image components in the factor loading matrix $\mathbf{\Lambda}$:

$$p(\mathbf{g}, \mathbf{y}) = N(\mathbf{g}; \boldsymbol{\mu} + \mathbf{\Lambda}\mathbf{y}, \mathbf{\Phi}) N(\mathbf{y}; \mathbf{0}, \mathbf{I}) \tag{1}$$

where $\mathbf{\Phi}$ constitutes the non-uniform image noise, i.e., the variability not captured in the subspace model. $\mathbf{\Lambda}$ is an $n \times k$ matrix used to expand from the k-dimensional subspace into a full n-dimensional one, where n is the number of pixels in the image $\mathbf{g}$. The parameters $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\boldsymbol{\mu}$ can be learned by maximizing the likelihood of a set of images $\{\mathbf{g}_t\}$,

$$\log p(\{\mathbf{g}_t\}) = \log \sum_t \int_{\mathbf{y}_t} p(\mathbf{g}_t, \mathbf{y}_t), \tag{2}$$

and a good low-dimensional representation of the image tends to be $E[\mathbf{y}|\mathbf{g}]$. The above probability model, called factor analysis (FA), also allows for the design of the optimal encoding strategy for the factors $\mathbf{y}$. A realted method, principal component analysis, was used by Moghaddam and

Pentland for face recognition and compression [32]. By limiting their representation to the central part of the face they were able to represent each image in a carefully manually preprocessed database, with only 85 bytes describing 100 face factors $\mathbf{y}$. In our case, we need a more robust coding scheme that does not require precise manual registration of images, and can encode more than just the central region of the face. We also include hair and the face shape, in order to lower the probability of false positive matches.

Recently, an extension of the subspace models that takes into account the possible transformation of the facial image, such as translations, rotations and scale has been proposed in [21]. In this model, called transformed component analysis (TCA), an additional random transformation variable T is applied to the image expanded from $\mathbf{y}$, and a new image $\mathbf{h}$ is observed:

$$p(\mathbf{h}, \mathbf{g}, \mathbf{y}) = N(\mathbf{h}; \mathbf{T}\mathbf{g}, \mathbf{\Psi}) N(\mathbf{g}; \boldsymbol{\mu} + \mathbf{\Lambda}\mathbf{y}, \mathbf{\Phi}) N(\mathbf{y}; \mathbf{0}, \mathbf{I}) p(\mathbf{T}).$$

Such a model, when trained on an image set tends to automatically align all images to create the most compact subspace representation. The regular subspace models, in presence of tranformational variability in the training data will tend to create blurry models, while TCA creates sharper components.

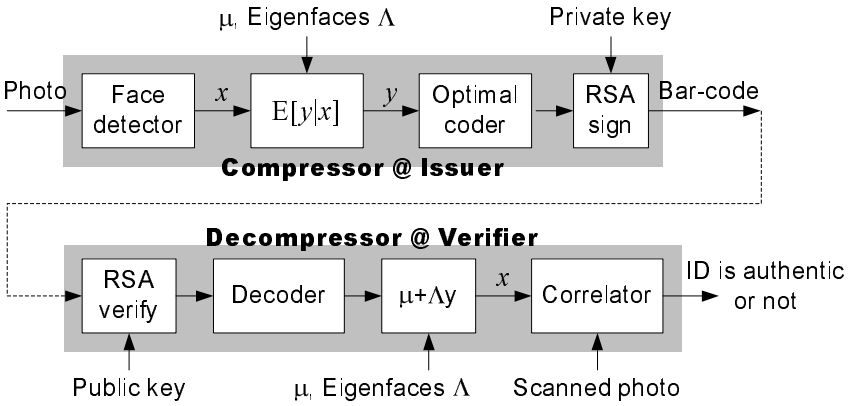

Figure 2: Block diagram of the face compression and decompression algorithm encapsulated within the FaceCert issuing and verification system. The $\mathbf{\Lambda}$-subspace model $\mathbf{y}$ follows a Gaussian distribution and thus can be encoded close to its rate-distortion limit.

A hierarchical generative model like this is naturally suited for efficient compression, as it decomposes the variability in the data. To develop the coder, the model is first trained on a large number of face images, i.e., the subspace origin $\boldsymbol{\mu}$ and subspace vectors $\mathbf{\Lambda}$ are estimated together with the pixel noise levels $\mathbf{\Phi}$ and distribution over the used transformations (rotations, scales, shifts and deformations) $p(\mathbf{T})$. Then, for a particular image to be encoded, the hidden variables are inferred and each of the conditional probability distributions, i.e., $p(\mathbf{T})$, $p(\mathbf{y})$, $p(\mathbf{g}|\mathbf{y})$, $p(\mathbf{h}|\mathbf{g}, \mathbf{T})$, is used in an appropriate entropy coder to create codewords for describing the geometric position and deformation of the image, as well as its subspace coordinates. As the model distributions are either multinomial or Gaussian, this procedure is straightforward. For example, for a Gaussian source a non-uniform quantization is used that is fine close to the mean of the Gaussian and coarse in the unlikely areas of the subspace.

The transformation information is then combined with the face cropping information needed to capture the face from the scanned ID and encoded in the barcode, while the subspace encoding is illustrated in Figure 2. First, given an ID photograph, we identify the facial structure to be modelled $\mathbf{x} = N(\mathbf{\Lambda}\mathbf{y} + \boldsymbol{\mu}, \mathbf{\Phi})$ with eigenfaces using a face detection algorithm [33], [34]. Vector $\boldsymbol{\mu}$

denotes the first order statistics of the input image $\mathbf{x}$. As the posterior $p(\mathbf{y}|\mathbf{x})$ can be computed using the Bayesian rule, hence we compute:

$$\begin{aligned}\log p(\mathbf{y}|\mathbf{x}) &= -\log p(\mathbf{x}) - \frac{\mathbf{1}}{\mathbf{2}}\mathbf{y}\mathbf{y}' - \frac{\mathbf{1}}{\mathbf{2}}\log(\mathbf{2}\pi\mathbf{I}) \\ &\quad -\frac{1}{2}(\mathbf{x}-\mathbf{\Lambda}\mathbf{y}-\boldsymbol{\mu})'\mathbf{\Phi}^{-\mathbf{1}}(\mathbf{x}-\mathbf{\Lambda}\mathbf{y}-\boldsymbol{\mu}) - \frac{\mathbf{1}}{\mathbf{2}}\log(\mathbf{2}\pi\mathbf{\Phi}) \qquad (3)\end{aligned}$$

which points to: $E[\mathbf{y}|\mathbf{x}] = \hat{\mathbf{y}} = (\mathbf{I} + \mathbf{\Lambda}'\mathbf{\Phi}^{-\mathbf{1}}\mathbf{\Lambda})^{-\mathbf{1}}\mathbf{\Lambda}'\mathbf{\Phi}^{-\mathbf{1}}(\mathbf{x}-\boldsymbol{\mu})$. Assuming $\mathbf{\Phi} = \sigma^2 I, \sigma \to 0$, we conclude that $E[\mathbf{y}|\mathbf{x}] = \hat{\mathbf{y}} = (\mathbf{\Lambda}'\mathbf{\Phi}^{-\mathbf{1}}\mathbf{\Lambda})^{-\mathbf{1}}\mathbf{\Lambda}'\mathbf{\Phi}^{-\mathbf{1}}(\mathbf{x}-\boldsymbol{\mu})$ which in the case when the basis vectors are orthogonal (e.g., $\mathbf{\Lambda}$ has been derived using PCA [20]) results in a simple least-squares approximation $\hat{\mathbf{y}} = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-\mathbf{1}}\mathbf{\Lambda}'(\mathbf{x}-\boldsymbol{\mu})$. In the $\mathbf{\Lambda}$-subspace, $\hat{\mathbf{y}}$ follows a Gaussian distribution, and thus can be efficiently encoded using codes with long block lengths (for analysis see [35], [36]), so as to approach the theoretical rate-distortion limit for the distribution illustrated in Figure 4.

3.1 Face Compression Illustration

We conducted several experiments in order to evaluate system performance. We trained $\mathbf{\Lambda}$ using 400 images of 64x64 faces extracted from personal photo collections of our colleagues employees using a face detection algorithm that follows the work of Viola et al. [34]. The resulting dataset contains alignment errors that were dealt with automatically by the transformed component analysis. We tested the coding performance on the Yale and Rockefeller face databases. Later in the paper we also report a separate field test of the entire creation and verification process by issued over 4000 IDs in two days and estimating the false positive and negative verification rates.

Figure 3: Five faces extracted from the Yale face database and the compressed images using JPEG (second row), PCA (third) and TCA (fourth). TCA achieved an RMSE of about ten intensity levels, considerably bellow the difference between any two images in the set. Both TCA and PCA were trained on a separate unrelated database of 400 images derived from personal digital photo collections.

In Figure 3 we show comparison between the JPEG, PCA and TCA coders on several faces in the test set. On average, at low bitrates, we were able to make JPEG encode the gray level images with 255 levels with 360 bytes and a root mean square error $rmse_{jpeg} = 36$, while both PCA and TCA performed better, with $rmse_{PCA} = 17$, $rmse_{TCA} = 10$, and with significantly

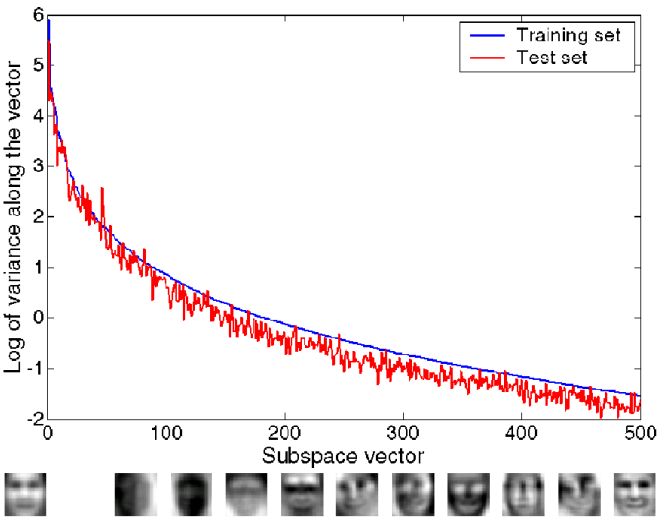

Figure 4: The distribution over the coordinates (strengths of the subspace vectors, or principal components) for the training set (blue), and a test set (red). According to the rate-distortion analysis of the blue distribution computed on the training set of 10000 images, for errors of roughly one intensity level out of 255, the image code would be only about 500 bits long. Bellow, we show the mean and the first ten subspace vectors.

lower bit rates of about 200 bytes for a 200-dimensional representation of images. TCA models used only shifts as the set of possible transformations $\mathbf{T}$. The *rmse* differences among the images in the test set were between 35 and 65, even for images of the same people with slightly different expressions. Thus, the TCA result is well beyond the error of random photo replacement.

Figure 4 shows in red the distribution of component strengths over the coordinates in the subspace. For this distribution, the optimal rate-distortion function indicates that for the error of standard deviation of 0.5 intensity levels (out of the 255), the number of bits needed to encode the image is about 500[3]. In other words, at 500 bits per face image, the coding error is expected to be smaller than 0.5% of the dynamic range of the image. This value is far bellow the scanning error of the system. On the same plot, in blue we plot the distribution over the subspace coordinates of images in a separate small face dataset (165 images), using the derived subspace vectors (first ten of which are shown at the bottom of the figure). Note again that this results depends on fine alignment that TCA algorithm provides. In practice, it is possible that to reduce the cost of creation a coarser alignment would be performed. In the field test we describe later, for example, we used the face code that was 1000 bits long.

4 FaceCerts - Verification

FaceCert verification consists of simple template matching. To be in accordance with the models in the previous section, a likelihood over the windows in the image can be used as a cost metric instead of template differences. For example, to use the likelihood as the similarity measure, one would take the message $\mathbf{f}$, extract the window size and detection threshold *thr* as

[3]Result reported for Yale database. Images in the Rockefeller database required about 1600 bits for similar performance.

well as the subspace parameters $\mathbf{y}$ to compute:

$$\log p(\mathbf{h}|\mathbf{y}) = \int_{\mathbf{T},\mathbf{g}} \log p(\mathbf{h}, \mathbf{g}, \mathbf{T}|\mathbf{y}), \tag{4}$$

for all windows of appropriate size. If $\max_{\mathbf{h}} \log p(\mathbf{h}) > thr$, then the ID does contain the face encoded in the bar code. If the position of the isolated face is stored in the barcode, the integration over transformation $\mathbf{T}$ is not necessary.

The detection threshold thr depends on the compression-print-scan error which FaceCerts must tolerate. This error can be used by the adversary to minimally modify the facial features on an image from a given collection of valid FaceCert IDs in order to create another image possibly as close as possible to adversary's facial features. In the next section of the paper we show that the combined compression and scanning error can be made so low that such an attack becomes futile - it leads to the altered face that is virtually indistinguishable from the original.

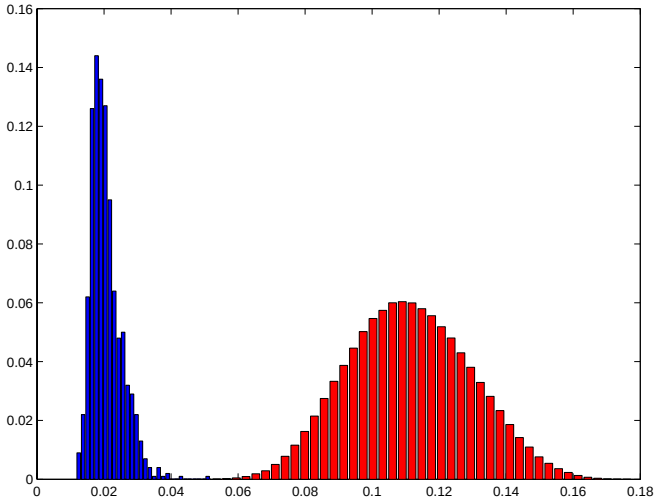

Figure 5: The distribution of the compression errors vs. the distribution over the pairwise distances on the set of 4239 faces we collected in our field test. The abscissa denotes the normalized Euclidean distance between two images, whereas the ordinate quantifies the distributions of interest.

5 Experiments: A Field Test

To test the entire solution, we developed and installed fully automated ID creation centers at a technology exhibition visited by thousands of people. The visitors created their own IDs by scanning their existing smartcard badges to provide personal information and then standing in front of the camera which took a snapshot of their face. Then, the face detection software localized the face allowing both proper framing of the photograph and the speedup in the face compression algorithm described above by reducing the search space for the transformation variable. The proper FaceCert ID was then printed on an inexpensive business card paper and issued to the user, who could then scan it at various stations equipped with business card readers and scattered across the exhibit floor. The entire print and scan test was thus performed without any manual intervention.

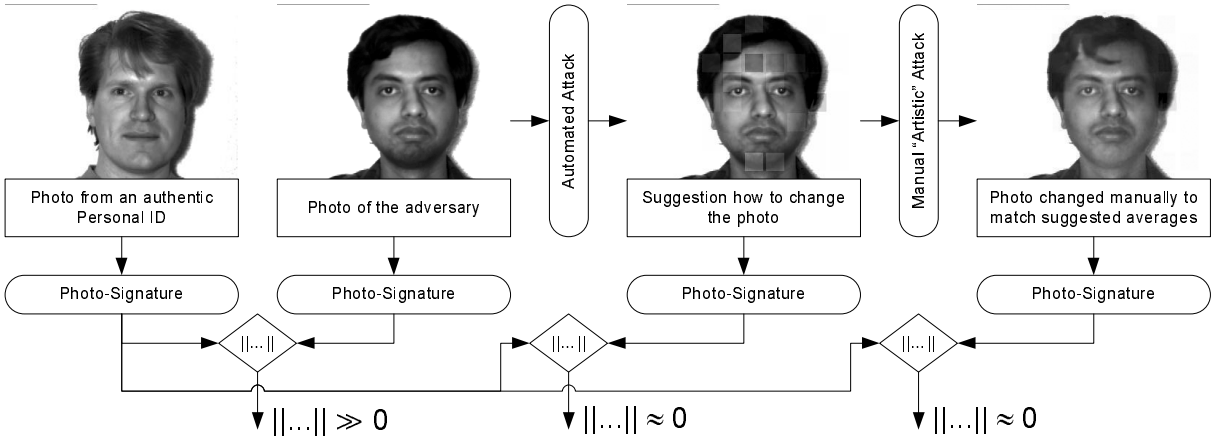

Figure 6: An example of the attack on the O'Gorman-Rabinovich Personal ID System. A photo from an authentic ID card (first photo) is obtained by the adversary (second photo) who does not resemble the person on the authentic ID card. Therefore, his photo has a different photo-signature. After applying the automated attack, the adversary obtains the set of guidelines on how the first order statistics of his photo must be changed (third photo) and finally, the adversary artistically edits his photo so that the suggested changes in the first order statistics are fulfilled while having a realistic visual appearance (fourth photo).

In this way we collected 4239 faces and corresponding FaceCerts. Before the test, we had estimated the best generative model parameters for the code length of 1000 bits using EM optimization on a database with only 1000 faces. The compression error typically achieved using this face model and data allocation, was within 2-3% of the dynamic range. The scanning error was about 1% of the range. Note however that the scanning error can be reduced by printer/scanner combo calibration to be virtually nonexistant. Also, the compression error can be reduced to sub-2% levels by increasing the code to 2000 bits.

For our target code length of 100 bits and the inexpensive combination of an off-the-shelf printer and a business card reader, we had set up the detection threshold to 5% of the dynamic range of the images. We anticipated that such a threshold was high enough not to expect any false negatives (failed scans of valid IDs), and so none of the FaceCerts that were properly created failed the verification step.

To get a sense of the probability of false positives (IDs with photograph replaced that still passed the verification), we computed all pairwise distances on the set of collected faces as well as all compression errors (see Figure 5) and plotted the probability distributions on a single graph. In our event with 4239 distinct people, the probability of face substitution passing the verification step, i.e., that one of the computed pairwise distances was bellow the threshold of 5% was in the order 10^{-6}. Note that this probability is significantly lower than the probability that one finds a look-a-like (estimated between $10^{-4} - 10^{-3}$). As we have stressed in Subsection 1.1, by using other biometric features within the FaceCert system, one can rectify the deficiency of using only person's face for identification. Finally, we looked at the handful image pairs that could be interchanged on the IDs based upon our threshold selection; they indeed looked sufficiently similar that a human verifier would not make the difference between them.[4]

Note that data allocation of 1000 bits leads to a barcode of the size equal to roughly one fifth of the size of the image on the ID. While our particular implementation only targeted a

[4]Unfortunately, due to privacy reasons we cannot show any photographs from the face database in the paper.

particular price/security ratio, there is plenty of ways to improve the rates depending on the application requirements. Techniques such as increasing in the barcode size, calibrating the scanner/printer combo, retraining the compressor on a larger database can be used to set the error rate to practically an arbitrary level without any significant change in the core technologies described in the paper.

6 Cryptanalysis of the O'Gorman-Rabinovich ID Card System

In order to demonstrate the effectiveness of our approach with respect to an existing technology, in this section, we briefly overview the key ingredient of the O'Gorman and Rabinovich personal ID system and then present a simple polynomial attack that guides an adversary to edit her photograph such that it has the same digital signature as a given photograph that is imprinted on an authentic ID card.

The ID card technology presented by O'Gorman and Rabinovich in [23] relies on a specific image digest function the authors call a photo signature method (PSM). PSM is a digest of the photographic content on a certified license. The objectives in PSM design are four-fold: (*i*) it must be a unique (or very close to a unique) identifier of the photograph; (*ii*) PSM must be concise (320 bits for a DSA signature); (*iii*) PSM must be invariant to noise, so that a PSM that has been subject to fading, dirt, nonuniform contrast change, and other common card noise still yields the same as or very close to the original PSM; (*iv*) the photograph must be difficult to modify, so as to match its PSM with that of a different photograph. As oppose to FaceCerts, both the original PSM as well as its cryptographic signature are printed on an ID card. Verification of the ID card is governed by deployed crypto-protocols.

The authors propose the following PSM: **1)** Transform the original image into lower resolution images by performing low-pass filtering and subsampling by successive powers of two. These images are of sizes: level $l_0 \rightarrow N \times N$; level $l_1 \rightarrow N/2 \times N/2$; level $l_2 \rightarrow N/4 \times N/4$; etc. **2)** Choose all or some of these multiresolution levels, $l_{L1} \leq l \leq l_{L2}$, and place a grid of size $I \times J$ on each. At each grid intersection location, determine the average of pixels in $k \times k$-sized pixel neighborhoods, $G^l(i,j), 0 \leq i < I, 0 \leq j < J$. **3)** For each grid point, a feature is determined. This feature represents the relative intensities of the neighboring grid points. The feature $S^l(i,j)$ is contained in 8 bits, where each bit corresponds to the eight grid neighbors and where a bit is one if the neighbor is greater than the grid value and zero otherwise. **4)** For PSM elements at grid intersection points of each level, l, represented as $S^l(i,j)$, each PSM element is the next grid intersection value from the next level. For instance, if $L1 = 1$ and $L2 = 4$, then the PSM elements are:

$$\begin{aligned} S(i,j) = \{&S^1(1,1), S^2(1,2), S^3(1,3), S^4(1,4), \\ &S^1(1,5), S^2(1,6), \ldots\} \end{aligned} \tag{5}$$

In general, for a total number of chosen levels, L, lowest chosen level, $L1$, and a grid size of $I \times J$, the sequence of levels is, $l(i,j) = [L1 + (i + jJ)]_L$.

Given a valid photo ID card and a photograph of the adversary, the goal of the attack is to edit adversary's photo such that its digital signature equals the one presented on the valid ID card. In that case, the adversary can replace the original photo on the ID card with her own. In this manuscript, we present an attack that achieves this goal by relying on two separate procedures: **a)** an automated phase – which creates a guidance for the next step, and **b)** artistic follow-up editing – which manipulates a photograph to satisfy the constraints posed by the

automated step a). The steps of the attack and their effect on the corresponding photographs is illustrated in Figure 6.

The goal of the automated step of the attack is to create a set of guidelines for artistic editing. The idea behind the guidelines is based upon the fact that the important statistic collected for a given image is based upon the averages (first order statistics) of certain image regions $G^l(i,j)$ and most importantly, not their values but relations between them (e.g., whether $G^l(i,j) < G^l(i+1,j)$). Therefore, the attack is aiming to change the first order statistics of adversary's image such that its image regions obey the same relations as the relations of the image on an authentic ID card, while inducing minimal change to the adversarial image. An example of such changes, marked as darkened or enlightened rectangles on the adversarial image after the automated attack (see Figure 6), would guide the artistic editor of the image on what has to be changed on the image such that it is semantically appealing but still satisfies the suggested first order statistics. An example in Figure 6 shows that only slight edit of person's hair, an enlightened chin, and darkened right ear, is sufficient to equalize the PSMs of photos of two persons who do not resemble each other[5].

The algorithm that creates the guidelines for artistic editing is presented using the following pseudo-code:

Lets denote the first order statistics of adversary's photo as $G_A^L(i,j)$, the authentic photo as $G_o^L(i,j)$, and the resulting photo as $G_R^L(i,j)$ for all considered levels $L1 \leq L \leq L2$.

for all $L1 \leq L \leq L2$, $0 \leq i < I$, and $0 \leq j < J$
  set $G_R^L(i,j) = G_o^L(i,j)$.
repeat $I \times J \times L$ times
  for each $L1 \leq L \leq L2$
    for each $0 \leq i < I$
      for each $0 \leq j < J$
        set $G_R^L(i,j)$ to a value as close as possible to $G_A^L(i,j)$ such that the relations of $G_R^L(i,j)$ with respect to its neighbors in the resulting image are the same as the relations of $G_o^L(i,j)$ and its neighbors in the original authentic image.
        update all $G_R^l(i',j')$ that intersect with $G_R^L(i,j)$.

The attack first sets the first order statistics of the resulting image G_R to the ones exhibited by the authentic image G_o. Then, it iteratively reduces the distances between individual components of G_R and G_A such that for each alteration in G_R, the $\gtrless$ relations between the altered grid component $G_R^L(i,j)$ and its neighbors stay the same as the relations between the corresponding $G_o^L(i,j)$ and its neighbors in the authentic image. By construction, this algorithm leads to a set of first order statistics G_R which is at minimal linear distance with respect to G_A and still satisfies all the constraints imposed by the photo-signature of G_o. The worst-case run-time of the algorithm is $\mathcal{O}(I^2J^2L^2)$; however, in all empirical runs of this attack, we have achieved the desired result in $\mathcal{O}(IJL)$. The outermost loop of the attack is aborted if no G_R component is changed throughout a single run of that loop. An example of how adversary's photograph looks like after it is updated for corresponding distances between the final G_R and G_A, is presented as second from the right in Figure 6. Finally, the adversary encounters a trivial and only artistically challenging task to edit his photograph so that the changes in the first order statistics have visually realistic semantics as presented in the rightmost photo in Figure 6.

[5]Note that the quality of edits is poor on the example image.

7 Conclusion

In this manuscript, we propose FaceCerts, a system for creating and verifying secure identity certificates. FaceCerts prevent tempering with the photograph or associated text by encoding the cryptographic signature of the face and text in a compact barcode readable by ordinary scanners.

Today, in a typical scenario, the verifier of the ID needs to connect to a remote database and retrieve a stored photograph for the comparison with the ID. In our system, all the necessary data for verification is securely stored on the ID itself, in a form of a barcode. The system does not depend on face recognition technology, but rather on the much more reliable face compression. We show that 100x66 pixel color face images can be compressed to about 1000 bits, while the color barcodes can reliably carry two to three thousand bits. A potential issuer of FaceCerts, such as a government agency, for example, may have a significantly larger database of facial images to train even better compression systems, using one of the methods we described.

References

[1] R.L. Van Renesse. Optical Document Security. Artech House, 1998.

[2] LaserCard Systems Corp. Details available from: `http://www.lasercard.com`.

[3] R.L. Rivest, A. Shamir, and L.A. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, vol.21, no.2, pp.120–6, 1978.

[4] A. Freier, P. Karlton, and P. Kocher. The SSL protocol Version 3. December 1995.

[5] T. Wang. Issues In Brief: The Debate over a National Identification Card. The Century Foundation, 2002.

[6] CASIA Iris Image Database. On-line presence at: `http://www.sinobiometrics.com`.

[7] A.J. Menezes, P.C. Van Oorschot, and S.A. Vanstone. Handbook of Applied Cryptography. CRC Press, 1996.

[8] D. Schonberg and D. Kirovski. Iris Compression for Cryptographically Secure Person Identification. *IEEE Data Compression Conference*, to appear, 2004.

[9] P.Y. Simard, D. Steinkraus, and J. Platt. Best Practice for Convolutional Neural Networks Applied to Visual Document Analysis. *IEEE International Conference on Document Analysis and Recogntion*, pp.958–962, 2003.

[10] Adobe Inc. OCR fonts. On-line presence at: `http://www.adobe.com/type/browser/P/P_058.html`.

[11] I.S. Reed and G. Solomon. Polynomial Codes over Certain Finite Fields. *SIAM Journal of Applied Mathematics*, pp.300–304, 1960.

[12] Wilhelm Research. Technical report available on-line at: `http://www.wilhelm-research.com`.

[13] Symbol Technologies Inc. The PDF417 Barcode. Details available from: `http://www.pdf417.com`.

[14] Y. Desmedt and A. Odlyzko. A Chosen Text Attack on the RSA Cryptosystem and Some Discrete Logarithm Schemes. *CRYPTO*, Springer-Verlag, pp.516–522, 1985.

[15] D. Bleichenbacher. Chosen Ciphertext Attacks Against Protocols Based on the RSA Encryption Standard PKCS #1. *CRYPTO*, Springer-Verlag, pp.1–12, 1998.

[16] J.-S. Coron, D. Naccache, and J.P. Stern. A New Signature Forgery Strategy. *CRYPTO*, Springer-Verlag, pp.1–18, 1999.

[17] M. Bellare and P. Rogaway. The exact security of digital signatures: how to sign with RSA and Rabin. *EUROCRYPT*, Springer-Verlag, pp.399–414, 1996.

[18] M. Bellare and P. Rogaway. Optimal Asymmetric Encryption How to Encrypt with RSA. *EUROCRYPT*, Springer-Verlag, pp.92–111, 1994.

[19] M.A. Turk and A.P. Pentland. Face Recognition Using Eigenfaces. *IEEE CVPR*, pp.586–91, 1991.

[20] I.T. Jolliffe. Principal Component Analysis. Springer-Verlag, 1986.

[21] B.J. Frey and N. Jojic. Transformed Component Analysis. *ICCV*, pp.1190–6, 1999.

[22] L.A. Ray and R.N. Ellson. Method and Apparatus for Credit Card Verification. U.S. Patent 5,321,751, June 1994.

[23] L. O'Gorman and I. Rabinovich. Secure identication documents via pattern recognition and public-key crypto. *PAMI*, pp.1097–102, 1998.

[24] Biometric Consortium. On-line presence at: `http://www.biometrics.org`.

[25] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of Artificial Gummy Fingers on Fingerprint Systems. *Optical Security and Counterfeit Deterrence Techniques IV*, SPIE, vol.4677, 2002.

[26] A.J. Mansfield and J.L. Wayman. Best Practices in testing and reporting performance of Biometric Devices. National Physical Laboratory, technical report available at:
`http://www.cesg.gov.uk/technology/biometrics`
`/media/Best%20Practice.pdf`.

[27] G.I. Davida, Y. Frankel, and B.J. Matt. On the relation of error correction and cryptography to an offline biometric based identication scheme. *Proceedings of the Workshop on Coding and Cryptography*, 1999.

[28] R. Anderson and M. Kuhn. Low cost attacks on tamper resistant devices. *International Workshop on Security Protocols*, vol.1361, pp.125–136, 1997.

[29] P.C. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. *CRYPTO*, Springer-Verlag, pp.388–397, 1999.

[30] J.-F. Dhem, F. Koeune, P.-A. Leroux, P. Mestre, J.-J. Quisquater, and J.-L. Willems. A Practical Implementation of the Timing Attack. *CARDIS*, pp.167–182, 1998.

[31] D. Kirovski, H. Malvar, and Y. Yacobi. A Dual Watermarking and Fingerprinting System. *ACM Multimedia*, 2002.

[32] B. Moghaddam, A. Pentland. Probabilistic visual learning for object representation. *Early Visual Learning*, pp.99–130, 1996.

[33] G.-D. Guo and H.-J. Zhang. Boosting for Fast Face Recognition. *Personal communication.*

[34] P. Viola et al. A unified framework for face datection and recogntion. *Learning workshop, Snowbird*, 2002.

[35] T.M. Cover and J.A. Thomas. Elements of Information Theory. John Wiley and Sons, Inc., 1991.

[36] A. Gersho and R. Gray. Quantization and Signal Compression. Kluwer, 1992.