Illumination Invariant Representation for Privacy Preserving Face Identification

Boaz Moskovich
Computer Science,University of Haifa
Carmel, Haifa 31905, Israel
boazmosk@gmail.com

Margarita Osadchy
Computer Science,University of Haifa
Carmel, Haifa 31905, Israel
rita@cs.haifa.ac.il

Abstract

Most effective face recognition methods store biometric information in the clear. Doing so exposes those systems to the risk of identity theft and violation of privacy. This problem significantly narrows the practical use of face recognition technology. Recent methods for privacy preserving face recognition address face verification task. Most of them are unable to generalize to unseen conditions and require a large number of images of every user for training. We address the problem of face identification, which is more useful in security applications, and propose a binary, illumination invariant representation that can be easily integrated with various efficient cryptographic tools for protection. We propose several privacy preserving applications for our representation and test it on a number of benchmark databases to show its robustness to severe illumination changes, occlusions, and some other appearance variations.

1. Introduction

Our goal is to build privacy preserving methods for face identification in security applications. Incorporating privacy protection in face recognition is a very important task, because once the recognition system storage is compromised, the stored data can be misused in different ways (see [38, 31] for discussion about the risks of using unprotected biometric systems). Privacy exposure is a major factor in public opposition to the use of face recognition in everyday life.

We consider two types of protection: 1) The storage of a database and the matching is done by the same machine. The database is encrypted and the matching is done on the encrypted data. 2) A recognition system is composed of two major components, a server, which stores a list of faces, and a client, which holds a representation of a single face. The system must find out if the face held by the client matches any of the faces in the server's list. The matching must be done in such a way that the server and client must not learn any information except whether there is a match between the client's input and a face in the server's list. Both applications protect the identity of people, but rely on different cryptographic tools.

Adding a protection mechanism to a face recognition algorithm is a challenging task because of two related problems: Two images of the same person are never identical. Representations used in recognition are therefore designed to produce the same results for similar inputs. In cryptographic algorithms (e.g., for password authentication) only identical inputs enable successful authentication. To bridge this gap, there have been attempts in cryptography to develop noise resistant one-way hash functions [21, 20, 37, 13, 7]. Namely, functions which are hard to invert but which provide the same image for inputs that are close to each other, and can therefore use "fuzzy" readings as keys. The second problem is that effective face recognition algorithms employ continuous representations which are compared by complex measures of similarity that in some cases are not even metric. Such representations are not supported by cryptographic algorithms, which are essentially discrete in nature. A naive conversion from the continuous comparison methods used in face recognition to a discrete measure like the Hamming distance, affects the accuracy of recognition. Thus we aim to design a binary face representation which is invariant to viewing conditions and can be matched by the Hamming distance. Such representations can be then easily integrated with efficient cryptographic protecting mechanisms. The details of the cryptographic tools are beyond the scope of this paper.

Face representation proposed in this paper is designed to be robust to illumination changes, occlusions, minor changes in facial expressions and it can be generalized to handle pose variation. Our experiments verify the robustness of the representation on different databases. However, the requirement that the biometric information cannot be stored in the clear is a very large obstacle for face recognition. Our representation satisfies this requirement, while minimizing degradation in recognition.

1.1. Related Work

Different aspects of face recognition have been addressed in numerous papers [26, 42]. Here we focus on the following subjects: illumination invariance, robustness to occlusions and local noise, and protection of the biometric data.

Face Recognition: An ideal face representation should be robust to viewing conditions, occlusions, and other variations of the image. Illumination variation is one of the major factors that greatly influence the appearance of human faces [1]. The majority of the illumination invariant methods take a holistic approach where an image of a face is represented as a vector in a high-dimensional space and the recognition is done on this vector. See [43, 39] for the overview of these methods. Holistic methods are robust to illumination changes, but when some parts of a face are occluded, most of them degrade significantly.

An alternative approach for face representation is based on the *part-based* paradigm, in which a face is represented by a collection of image fragments corresponding to different components of a face and their relative positions. Given a novel image, the recognition score is computed as a combination of recognition scores of every part in the representation and their geometric consistency. The robustness to viewing conditions is achieved by the use of robust image descriptors and by allowing some degree of deformation in the spatial configuration of parts. Part-based methods (e.g, [40, 24, 25, 5, 18, 44, 29, 28, 3]) are more robust to occlusions, deformations, and local noises than the holistic methods. If patches are small enough they are likely to correspond to planar patches on a 3D surface. For planar surfaces, normalized correlation could handle all of the variation due to illumination [8]. However, the planarity assumption doesn't always hold and in these cases the effects of illumination are much more complex and cannot be removed by normalization. Thus the robustness of the existing part-based methods to illumination variation is more limited than of the holistic methods.

The reliance of all these methods on real-valued representations makes it hard to apply cryptographic tools, which are inherently based on finite fields.

Protected biometrics: It is known that different acquisitions of the biometric data from the same person are never identical. Thus representations used in face recognition are not directly compatible with cryptographic mechanisms. To solve this problem different directions have been taken in both vision and cryptographic fields.

Fuzzy schemes: To solve the problem of fuzziness in biometric data, noise resistant one-way hash functions, referred to as fuzzy hashing (commitments), or as secure sketches and fuzzy extractors have been developed (e.g., [21, 20, 37, 13, 7]. These are functions F which map an input x to a sketch (or helper data) s while satisfying two properties: (1) *recovery*, namely that for any x' whose distance from x under some metric (which is limited to Hamming distance, set difference, and edit distance) is small, it is possible to efficiently recover x given x' and s; and (2) *security*, namely that it is hard to learn meaningful data about x from the sketch s. Fuzzy schemes assume that a real-valued biometric template is transformed into a fixed-length binary representation. Such an assumption requires methods that can perform the transformation without damaging the discriminative power of the original real-valued representation. The basic approach for this has been extracting real-valued features from a face image, applying quantization and coding of each feature, and concatenation of the output codes. The final decision has been made based on the Hamming distance between the binary strings of the stored and novel representations, because it is one of the few metrics supported by the fuzzy commitment schemes. Various combinations of image features and quintizers have been proposed for fuzzy schemes [23, 9, 11, 12, 10] (see [12] for more details). These methods focus on the verification task, which allows using many images for registration in order to cope with intra user variation. Inability of these algorithms to generalize to unseen conditions limits their usage. The extra step of quantization further damages the recognition strength of these methods.

Secure computation of face recognition: Here the biometric data is stored at the remote server, which is assumed to be much more secure than users' machines, and probe images are acquired by a client, which might be a camera taking photos of passersby. During recognition the server and client must not learn any information except for whether there is a match between the client's input and a face in the server's list. First attempts of conducting secure computation of face recognition [14, 32] employed the Eigenface method [36], which is unable to generalize to unseen conditions. The secure computation of matching, described in these papers, requires large network traffic and large memory which makes those systems not practical. Much more efficient protocols have been proposed for the secure computation of the Hamming distance [19]. Thus building a robust face representation that can be matched using the Hamming distance will greatly advance this area.

There are other methods in protected biometrics which are less related to our approach. Among them e.g, [31] that applies noninvertible distortion transform on the registration and verification copies of user's biometrics; [33] uses a user key to distort the input templates such that the peak of correlation corresponds to the same user. In [6], the approach is to transform a standard face representation (e.g., PCA, LDA) to a user specific window, in which robust distance is used for matching. The location of the window is encrypted, while the parameters of the transform and the offset inside the window is used as a helper data. They

report very impressive recognition results on frontal faces with mild changes in illumination and appearance. However, it's unclear how this approach will scale up to large variation in viewing conditions. Integrating it with more advanced methods that are invariant to viewing conditions seems hard, since these method require storing unprotected 3D face models or a number of images of the user under varying conditions for illumination compensation during the recognition phase. More detailed discussion on the revocable biometrics can be found in [31, 6].

2. Overview of the approach

In order to use privacy protecting cryptographic tools we must make the representation matchable by the Hamming distance (or the set difference), meaning that we need to find a way of converting a face image into a binary representation (or a set of indices). Previous methods took a discriminative approach in which a distribution of a particular face is compared against the distribution of all faces (which is a public data) to find the most distinguishable intervals, that are expressed as indices. Such an approach however does not scale up to variation of viewing conditions, since it can only handle cases, which are presented in the distribution. In this paper, we propose a generative approach, in which the face of a person, whose identity should be protected, is composed of fragments, obtained from a ***public*** set of faces. More explicitly, we use the facial composite principle (aka photo-robot) for face representation, which is based on the idea that facial features have a number of typical appearances and almost every face can be generated by combining fragments from ***other*** people. Although global similarity between faces is rare, the similarity of facial features or their fragments among different people is strong enough for recognition. Such an approach allows to apply complex, robust algorithms, that use real-valued representations in order to match facial features of a person, whose identity must be protected, to the facial features extracted from the public face set. If we associate each fragment in the public database with an unique index, then the choices, obtained by robust matching, can be expressed as indices into the public set. The resulting index-based representations can be compared by set difference or can be easily transformed to a binary representation (the details are given below) and matched using the Hamming distance.

More formally, let X denote a set of people enrolled in the recognition system. Assume that we have a database Y of faces unrelated to X. We make Y public, but we want to protect X. We define N salient parts using a regular grid around the textured areas of the face with each part having appearance and spatial component. Then for each part we build vocabularies of typical appearances (that will be called *words*) and vocabularies of quantized distances from the center of a face. The vocabularies are learned from Y – people who are ***unrelated*** to those registered in the system. In total we have N appearance vocabularies of size M and N spatial vocabularies of size Z. Each word in the vocabularies is associated with a bit in the binary representation of a face. This results in the representation of size $N(M + Z)$. K words that best match the appearance of the part in a given face are set to one, and the rest of the words are set to zero (similar process is done for spatial words). The resulting binary strings are matched using the Hamming distance and thus can be either incorporated with a cryptographic fuzzy schemes or matched using the secure Hamming distance computation [19].

A photo-robot representation works for humans, who can generalize over variation in viewing conditions. Using same principle in an automated system requires designing a mechanism that will cancel the effects of viewing conditions. We use similarity of face 3D shapes in order to estimate the viewing conditions in a given image and render part vocabularies to match the lighting in the image. Our construction represents parts by unordered set of appearances which contributes to the robustness against occlusions, shadows, highlights, and other local changes in appearance.

We preserve high entropy of the representation by using many patches with small overlaps between them. Choosing several matches per patch in the representation not only increases the robustness of the representation but also adds entropy (if the vocabularies are large enough).

In order to prevent cross-matching between databases, different vocabularies could be used. The difference can be either in patch locations or in individuals that are used to form the vocabularies.

Our representation requires only a single image of a person, which makes it very attractive for usage in settings, where there is no cooperation by the targeted persons.

Unlike other protected biometric methods, the proposed representation is inherently binary, thus there is no loss in recognition performance due to binarization. However some performance loss appears due to inability of cryptographic tools to efficiently find minimum in a privacy preserving manner. Thus in order to identify a probe image, instead of choosing the identity whose gallery image is closest to the probe image, we are required to apply series of verifications against every person in the gallery. This requires setting individual thresholds which are difficult to estimate using a single image for training.

3. Vocabularies Construction

Most local image descriptors cope with illumination by normalization and thus can cancel only an additive and multiplicative effects on image intensities. Descriptors that use gradient directions (e.g., SIFT [27]) should be also robust to changes in lighting direction. However in practice, it was

shown [44] that using SIFT in face representation helps with the pose more than with illumination. Thus we use 3D models of faces from the public set Y in order to render vocabularies with illumination that matches the input image.

3.1. Illumination Model

Given an image I of a person, whose identity must be protected, we use 3D models and textures of faces from the public set Y to estimate the lighting in I and render synthetic images of subjects from Y with the estimated lighting. The synthetic images are used to form vocabularies of parts with illumination similar to I. The estimation of lighting and image rendering is done using the model proposed by Basri and Jacobs [4]. This model approximates the set of images produced by a Lambertian object under varying illumination by a 9D linear subspace that is computed from the 3D model. The dimensions of this subspace are low-degree polynomial functions that are spherical harmonics of the surface normals of the face in a specific position, scaled by albedo, ie.:

$$b_{nm}(x,y) = \lambda(x,y) h_{nm}(\theta(x,y), \phi(x,y)), \qquad (1)$$

where b_{nm} are basis images of the person for ($0 \leq n \leq 2$, $-n \leq m \leq n$), $\lambda(x,y)$ is the albedo at pixel (x,y) and h_{nm} is the spherical harmonic evaluated at the surface normal (θ, ϕ) corresponding to pixel (x,y). Given an image I, [4] seeks a vector a that minimizes $\|Ba - I\|$. B denotes the basis images, arranged as a $k \times 9$ matrix, where k is the number of points in the image. Every column of B contains one harmonic image b_{nm}, as per equation 1. After solving this linear minimization problem, the low frequency components of the lighting can be derived from the coefficients a. In our settings instead of using 3D models of the people we want to protect, we approximate lighting using a 3D model of people from the public set. Due to shape similarity, such approximation is feasible, and has been actually used for 3D reconstruction ([22, 17]).

3.2. Offline Stage

During the offline stage we first compute the basis images B_i of every person in Y (Section 3.1). These are used later for lighting estimation and rendering vocabularies.

Next, we find the location of parts in every person in Y. Let I_Y denote an image of a person from Y and I_a denote an image of an average person, both rendered with frontal illumination. For each part in the representation we take a corresponding patch from I_a and search for its best match in I_Y. The patches are represented by SIFT and compared using $L2$ norm. The search is reduced to a window, the center of which is found by applying a linear transformation from the center of the patch in I_a to I_Y. The transformation is found using 5 correspondence points between I_a and I_Y. The same process is done for every person in Y.

To quantize similar appearances of parts we apply clustering on patches associated with the same part and take one representative per cluster (closest to the center) to be a word in the part vocabulary. Individuals that are not associated with any word are removed from Y. We end up with N vocabularies of parts. Each vocabulary contains M words and for each word it stores a triplet (i,x,y), where i is an index of the person it was extracted from; (x,y) are the coordinates of the word in the image of the person i. Since the vocabularies are constructed from Y, they stay fixed whether X changes or not, and thus no retraining is needed when a new subject is added to the system.

3.3. Adaptive Vocabularies

Given an image I that must be protected, we align it with every model in Y. Then, for each person i from Y we estimate lighting coefficients a_i by minimizing $\|B_i a_i - I\|$ (B_i are the basis images of person i) and render a synthetic image $J_i = B_i a_i$ of that person[1]. The vocabularies of visual words are generated by extracting fragments from the set of images $\{J_j\}_{j=1}^{|Y|}$ using the triplets (i,x,y) that were assigned for every word during the offline stage.

3.4. Spatial Vocabularies

The spatial information is modelled by the distance from the center of a part to the center of the face. During the offline stage, we estimate the distance distribution of each part and quantize it into Z bins, forming N spatial vocabularies of size Z. The estimation is done on the subjects from the public set Y.

4. Face Representation

Given an image I, we set the initial locations for the parts in I by applying a linear transformation on the centers of the corresponding parts in the image of the average model. The transformation is found using 5 correspondence points between the image of an average model and I. These points can be detected using automatic methods for locating facial features (e.g., [41, 16]).

A face representation is defined as a binary string s, which is composed of appearance component s^a and the spatial component s^s. Both appearance and spatial components are obtained by concatenation of N binary strings with one string per part: $s^a = [s_1^a, \ldots, s_N^a]$, for appearance component and $s^s = [s_1^s, \ldots, s_N^s]$ for spatial component.

Construction of s_i^a: We define a window centered at the initial position of part i in I. Then each word from the appearance vocabulary of part i is searched for in the defined window and its best match is noted. Let J_i^a denote the set of

[1] We do not use average model for light estimation, because light estimation is affected by skin tone (the spherical harmonics basis DC image).

indices that correspond to K words with the highest scores. A binary string s_i^a is defined as follows:

$$s_i^a(j) = \begin{cases} 1 & \text{if } j \in J_i^a \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $j = 1, .., M$ and M is the number of words in the vocabulary of part i.

Construction of s_i^s: Let d_i denote the distance between the initial location of the i-st part and the center of the face in I. A binary string s_i^s has the size Z and it's defined as follows:

$$s_i^s(j) = \begin{cases} 1 & \text{if } j \in J_i^s \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where J_i^s are the indexes of H quantized distances from the spatial vocabulary of part i that are closest to d_i.

Implementation Details: In the current implementation we use vocabularies with $M = 20$ visual words and $Z = 10$ quantized distance bins, and the number of face patches is $N = 30$. We empirically found that it is best to set the number of non zero bits in the appearance component of a part s_i^a to 4 (out of 20) and in the spatial component of a part s_s^a to 2 (out of 10).

5. Recognition

The proposed representation can be used with the fuzzy cryptographic schemes or it can be matched using secure Hamming distance computation [19]. Due to the space limit we will show the integration of our representation with only one protection mechanism, namely fuzzy extractors [13], but other integrations are also very easy. We choose fuzzy extractors over secure sketches because they provide not only error tolerance but also address the non-uniformity of the input. Our representation provides robust binarization of the face template, but it doesn't guarantee uniform distribution over the representation space.

5.1. Authentication using Fuzzy Extractors

Given a vector $\mathbf{s}$, the fuzzy extractor generates a random key R and a public recovery string P which leaks only negligible information about $\mathbf{s}$. Given a vector $\mathbf{s}\prime$ such that $\mathrm{dist}_{\mathrm{Hamming}}(\mathbf{s}\prime, \mathbf{s}) < t$, the recovery string P can recover R, without storing it in the system. The recovery string is able to correct up to t errors, with the upper bound on t is half of the length of $\mathbf{s}$. Fuzzy extractors can be used to protect biometric data as follows: During the enrollment stage, the system receives one image of a person, and computes a representation vector $\mathbf{s}$. To hide the representation a fuzzy extractor is applied (see [13] for details) and the generated tuple $\langle P_i, f(R_i)\rangle$ is stored for every person i (f is a one-way hash function). During the verification of a user i, the system receives a fresh image of the user, it generates a representation vector $\mathbf{s}'$ and tries to recover the key by applying the corresponding recovery string P_i. If the recovery from $\mathbf{s}'$ fails, the probe image is rejected for the identity i, otherwise $f(R_i')$ is compared with $f(R_i)$. If they are equal, the probe image is authenticated with the identity i.

Identification: Fuzzy schemes have been used in verification tasks, in which the matching of a probe is done against a single user. If the probe passes the user's threshold (which defines the upper bound on the number of errors that the recovery string will correct), then the system accepts the probe. In identification, matching is done against all the users in the gallery and the one that produces the best score is chosen as a match. The fuzzy protection schemes rely on user's threshold and thus cannot be directly applied to find the best score in a protected way. We suggest to implement identification as series of verifications against every person in the gallery. This way instead of best match we get a set of candidates for whom the probe passed their individual threshold. If the set contains only a single person then the proposed procedure is identical to the identification, otherwise the identification is ambiguous. To solve this problem we use an adaptive threshold. For each enrolled person we create a set of recovery strings, each capable of fixing different number of errors, starting from the individual threshold and decreasing by fixed step. The algorithm proceeds by applying recovery strings with decreasing error correction tolerance on $\mathbf{s}'$ until it maps to a single entity or no entity. In the latter case the probe is rejected. Such construction reveals more information then the original fuzzy extractors. However, if the thresholds are accurate enough, the number of recovery strings required to resolve the ambiguity is small. In our experiments for achieving 99.48% recognition rate on the CMU-PIE illumination subset(see Section 6) we needed on average 1.57 recovery strings per person. Thus the proposed construction is still acceptable in terms of privacy.

Suspects Detection: We consider an application in which a gallery is much smaller than the probe and the task is to detect whether the probe image matches one of the faces in the gallery. Such setting is very useful in surveillance applications in which the gallery could be a confidential list of criminals, terrorist or missing people and the probes are images streaming from public places. Here we can use only a single recovery string per person in the gallery. First, because the list is usually not very long and the probability that a probe will pass more than one individual threshold is relatively low. Second, even if it does, there is no harm, because if a person in the probe image matches any subject from the list, he or she should be turned to the authorities.

6. Experiments

We tested the proposed binary representation on several databases, checking its robustness to various factors that influence the intra-user variation. The tests were conducted on the benchmark data sets to allow comparison with the existing methods.

We tested the proposed representation in two types of applications: 1) *identification in a closed universe*, meaning that all probes are in the gallery (see [30]) and 2) *suspects detection*, where the gallery is much smaller than the probe and the task is to detect whether the probe image matches one of the faces in the gallery.

We constructed the public set Y from a sub-set of 34 3D models of faces supplied by USF[2].

6.1. Identification Test

Following FERET [30] evaluation of identification in a closed universe, we report the performance statistics by a graph of cumulative match. The horizontal axis of the graph is rank and the vertical axis is the probability of identification. In the large gallery test, FERET reports the top 50 matches for each probe (see further details about the evaluation protocol in [30]). Since other benchmark data sets include much less subjects, we set the number of top matches relative to the number of subjects in the gallery.

Large illumination Variation: We tested the robustness of the proposed binary representation to large illumination changes on the frontal pose subset of CMU-PIE database [35] and on the frontal pose subset of the Yale Face Database B and its extension[3] [15]. CMU-PIE contains images of 68 people under 43 illuminations. Half of the illuminations include ambient light and the other half doesn't. The images with ambient light include many subjects wearing glasses, and the subset without ambient light includes images with difficult illuminations in which half of the face is in shadows. The Yale Face Database B and the Extended Yale Face Database B [15] (the frontal subset) together include images of 38 people under 64 illuminations. Many of the illumination conditions in this set are extremely challenging.

For the CMU-PIE set we used frontal illumination with ambient light as a gallery image and other 42 illuminations as probes, in total 2856 probe images. Note that our setting tests not only illumination variation, but effects of glasses, since 28 out of 68 subjects wear glasses in the gallery image and then remove them in the half of the probe images (without ambient light). The results are shown in Figure 2. A separate test in which all galleries and probes include ambient light shows 100% recognition. A test in which all galleries and probes have no ambient light and no glasses shows 99%.

We formed another test set by combining The Yale Face Database B and extended Yale B together. We used frontal illumination as a gallery and the subsets 1 to 4 as probes ([15]). The results are reported in Figure 2 [4]

Near-frontal changes in pose, mild facial expressions, and illumination changes: Although current implementation of the system doesn't allow large variation in pose or facial expression, it can still handle some variation in these factors. To test our representation in more realistic setting, namely, near-frontal variation in pose and mild changes in facial expressions and illumination, we ran our system on the gallery and **fc** probe sets of FERET [30]; and on a subset of FRAV2D [34]. The gallery of FERET contains 1196 subjects and the **fc** probe set includes images of 194 subjects taken with a different camera and under different illumination. Besides the illumination, some variation in facial expression (smiling and blinking) and near-frontal pose variation is present between the gallery and the probe sets. FRAV2D set contains image of 109 subjects with variations in pose, facial expressions, illumination, and large occlusions. We conducted our tests on the subset of 16 images per subject, 12 of which include near frontal pose variations and mild changes in facial expression, and the other 4 include also variation in illumination, in total 1744 images. The results are reported in Figure 3. The recognition results on FERET are lower than in other tests, but are comparable with previously reported results [30, 2]. The gallery in this test is at least 11 times larger than in other sets, which explains the degradation in performance.

Illumination variation and Local Occlusions: We tested the effects of partial occlusions in eye, nose and mouth areas separately. To simulate occlusions we used a square area of the size of 20% of the face width filled with random noise (Figure 1 shows some examples of partial occlusions used in the test). Occlusion was applied to probe images only. Table 1 summarizes the recognition results, tested on the subset of CMU-PIE that includes images of 68 persons under 10 illuminations with ambient light and partial occlusions. The right column of the table describes the number of patches with at least half of their area occluded.

FRAV2D set [34] contains images with real occlusions, in which half of each face is occluded with a hand. We didn't test our representation on this set, because with such global occlusion at least half of the patches are irrelevant, which makes it impossible to use fuzzy extractors for protection (see Section 5).

[2] USF HumanID 3D Face Database, Courtesy of Prof. Sudeep Sarkar, University of South Florida, Tampa, FL

[3] http://vision.ucsd.edu/ leekc/ExtYaleDatabase/ExtYaleB.html

[4] Some previous papers on illumination invariance report results in the form of classification rate, which corresponds to the cumulative matching score for rank=1.

Figure 1. Examples of tested occlusions.

Occluded Part	rank=1 score	Num. of occluded parts
Left eye	95.37%	10/30
Mouth	99%	4/30
Nose	100%	3/30

Table 1. Partial occlusion results on a subset of CMU-PIE Database, containing images of all people in frontal pose and 10 illuminations

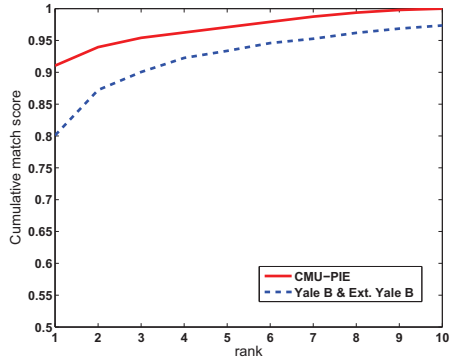

Figure 2. Identification test on test sets including dramatic illumination changes

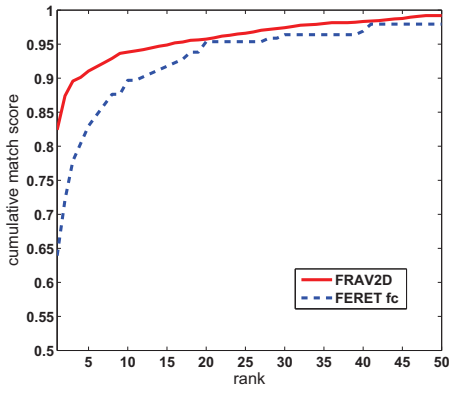

Figure 3. Identification test on test sets including near frontal pose change, mild change in facial expressions (smile, blink), and mild illumination changes.

6.2. Suspects Detection

This test simulates a real security systems which holds a confidential list of subjects, and detects whether a face from a stream of images coming from a public place matches one of the faces on the list. We use our representations followed by fuzzy extractors to hide the biometric information of the people on the confidential list. A single recovery string per person is computed to be used with the fuzzy extractors. These recovery strings correspond to the individual thresholds on the Hamming distance. To learn an individual threshold of subject i we construct an ensemble of people which includes other individuals from the list and images of unrelated people which represent typical inputs to the system (those can be easily obtained by a "dry run" of the camera before installing the recognition system.) An individual threshold for the ith subject is set based on the smallest Hamming distance between him and the rest of the people in the ensemble. When the system receives a probe face, it applies the verification procedure with every person on the list using the corresponding recovery string as described in Section 5.1. If the probe face passes the verification with at least one of the subjects on the list it is identified as a match. We tested the system on the frontal pose subset of CMU-PIE and on the FERET databases.

Similar application was addressed in [14, 32] and implemented as a secure computation of Eigenface [36]. Thus we also show the performance of the standard Eigenface (for which the images were cropped and normalized for brightness) on both sets of images.

CMU-PIE: The confidential list included 12 persons under frontal illumination. The test set contained 2912 images of 68 subjects, from which 504 belonged to the subjects from the list. All test images contained faces in a frontal pose under large illumination changes. About third of the subjects on the list wear glasses, but then remove them in half of the test images. The results are shown in Figure 4.

FERET: We took 100 subjects from the **fc** probe set for the confidential list, and used all 1196 gallery images (from which 100 belong to the subjects on the list) as a test set to simulate the stream of images. In this experiment the images on the list were taken with a different camera and with different viewing conditions than the test images. Figure 4 shows the results.

7. Conclusions

In this work, we presented illumination invariant, binary face representation which can be easily integrated with various existing cryptographic tools. The proposed approach showed very good performance in varying illumination conditions and robustness to partial occlusions. In the future work we plan to extend the proposed approach to pose variation and more significant changes in facial expression.

References

[1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. *PAMI*, 19(7):721–732, 1997.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *PAMI*, 28(12):2037–2041, 2006.

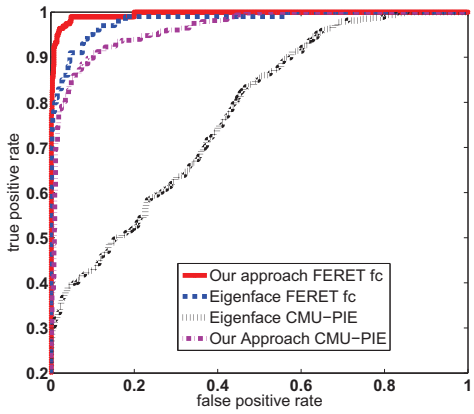

Figure 4. Suspects Detection test on CMU-PIE and FERET **fc** (This plot is best viewed in color)

[3] E. Bart, E. Byvatov, and S. Ullman. View-invariant recognition using corresponding object fragments. In *Proceedings of the European Conference on Computer Vision, Part II*, pages 152–165, 2004.

[4] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, 2003.

[5] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of sift features for face authentication. In *Proc. of IEEE Int Workshop on Biometrics, in association with CVPR*, 2006.

[6] T. Boult. Robust distance measures for face-recognition supporting revocable biometric tokens. In *FGR '06*, pages 560–566, 2006.

[7] X. Boyen, a. J. K. Y. Dodis, R. Ostrovsky, and A. Smith. Secure remote authentication using biometric data. In *Eurocrypt*, May 2005.

[8] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *PAMI*, 15(10):1042–1052, 1993.

[9] Y. Chang, W. Zhang, and T. Chen. Biometrics-based cryptographic key generation. In *ICME*, pages 2203–2206, 2004.

[10] B. Chen and V. Chandran. Biometric based cryptographic key generation from faces. *Digital Image Computing: Techniques and Applications*, 0, 2007.

[11] C. Chen, R. Veldhuis, T. Kevenaar, and A. Akkermans. Multi-bits biometric string generation based on the likelihood ratio. In *IEEE Conf. on Biometrics: Theory, Applications and Systems*, 2007.

[12] C. Chen, R. Veldhuis, T. Kevenaar, and A. Akkermans. Biometric binary string generation with detection rate optimized bit allocation. In *CVPR Workshop on Biometrics*, pages 1–7, 2008.

[13] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *Eurocrypt*, 2004.

[14] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft. Privacy-preserving face recognition. In *Privacy Enhancing Technologies (PET09)*, volume 5672 of LNCS, page 235253, 2009.

[15] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 23(6):643–660, 2001.

[16] N. Gourier, D. Hall, and J. L. Crowley. Facial features detection robust to pose, illumination and identity. In *Int'l Conf. on Systems Man and Cybernetics*, 2004.

[17] T. Hassner and R. Basri. Example based 3d reconstruction from single 2d images. In *CVPR Workshop Beyond Patches*, 2006.

[18] B. Heisele, T. Serre, and T. Poggio. A component-based framework for face detection and identification. *IJCV*, 74(2):167–181, 2007.

[19] A. Jarrous and B. Pinkas. Secure hamming distance based computation and its applications. In *(ACNS)*, 2009.

[20] A. Juels and M. Sudan. A fuzzy vault scheme. In *Symposium on Information Theory*, 2002.

[21] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *ACM Computers and Communication Security conference*, 1999.

[22] I. Kemelmacher and R. Basri. Molding face shapes by example. In *ECCV*, volume I, pages 277–288, 2006.

[23] T. Kevenaar, G. Schrijen, M. Veen, A. Akkermans, and F. Zuo. Face recognition with renewable and privacy preserving binary templates. In *IEEE Workshop on Automatic Identification Advanced Technologies*, pages 21–26, 2005.

[24] M. Lades, J. C. Vortbrüggen, J. Buhmann, J. Lange, R. P. W. C. von der Malsburg, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42:300–311, 1993.

[25] F. Li and H. Wechsler. Robust part-based face recognition using boosting and transduction. In *Biometrics: Theory, Applications, and Systems.*, pages 1–5, 2007.

[26] S. Z. Li and A. K. Jain, editors. *Face Recognition Across Pose and Illumination*. Springer-Verlag, June 2004.

[27] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[28] S. Lucey and T. Chen. Learning patch dependencies for improved pose mismatched face verification. In *CVPR*, 2006.

[29] Y. T. E. L. S. K. M. B. L. Luo, J. Ma. Person-specific sift features for face recognition. In *ICASSP 2007*, pages II–593–II–596, 2007.

[30] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *PAMI*, 22(10):1090–1104, 2000.

[31] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle. Generating cancelable fingerprint templates. *PAMI*, 29(4):561–572, 2007.

[32] A. Sadeghi, T. Schneider, and I. Wehrenberg. Efficient privacy-preserving face recognition. In *(ICISC 09)LNCS. Springer*, 2009.

[33] M. Savvides, B. V. K. V. Kumar, and P. K. Khosla. Cancelable biometric filters for face recognition. *ICVPR*, 3, 2004.

[34] A. Serrano, I. M. de Diego, C.Conde, E. Cabello, L. Shen, and L. Bai. Influence of wavelet frequency and orientation in an svm-based parallel gabor pca face verification system. In *IDEAL*, pages 219–228, 2007.

[35] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *PAMI*, 25:1615–1618, 2003.

[36] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[37] P. Tuyls and J. Goseling. Capacity and examples of template-protecting biometric authentication systems. In *ECCV Workshop BioAW*, 2004.

[38] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain. Biometric cryptosystems: Issues and challenges. *Proceedings of the IEEE*, 92(6):948–960, 2004.

[39] Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face re-lighting from a single image under harsh lighting conditions. In *CVPR*, volume 1, pages 1–8, 2007.

[40] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *PAMI*, 19(7):775–779, 1997.

[41] C. T. Yuen, M. Rizon, W. S. San, and M. Sugisaka. Automatic detection of face and facial features. In *ISPRA'08*, pages 230–234, 2008.

[42] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.

[43] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. Jacobs. Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *PAMI*, 29(2):230–245, 2007.

[44] Z. Zhou, A. Ganesh, J. Wright, S. Tsai, and Y. Ma. Nearest-subspace patch matching for face recognition under varying pose and illumination. In *FG*, pages 1–8, 2008.