Unconstrained Periocular Biometric Acquisition and Recognition Using COTS PTZ Camera for Uncooperative and Non-Cooperative Subjects

Felix Juefei-Xu and Marios Savvides
CyLab Biometrics Center, Electrical and Computer Engineering
Carnegie Mellon University
Felixu@cmu.edu, msavvid@ri.cmu.edu

Abstract

We propose an acquisition and recognition system based only on periocular biometric using the COTS PTZ camera to tackle the difficulty that the full face recognition approach has encountered in highly unconstrained real-world scenario, especially for capturing and recognizing uncooperative and non-cooperative subjects with expression, closed eyes, and facial occlusions. We evaluate our algorithm on the periocular region and compare that to the performance of the full face on the **Compass** *database we have collected. The results have shown that the periocular region, when tackling unconstrained matching, is a much better choice than the full face for face recognition even with less than 2/5 the size of the full face. To be more specific, the periocular matching across all facial manners,* i.e.*, neutral expression, smiling expression, closed eyes, and facial occlusion, is able to achieve 60.7% verification rate at 0.1% false accept rate, a 16.9% performance boost over the full face.*

1. Introduction

In this work, we present a novel acquisition and recognition system using the commercial off-the-shelf (COTS) pan-tilt-zoom (PTZ) camera. To be more specific, we consider the periocular biometrics under highly unconstrained environment and perform the recognition task for uncooperative and non-cooperative subjects. Uncooperative subject is an individual who actively tries to deny the capture of his/her biometric data, while non-cooperative subject is an individual who is not aware that his/her biometric sample is being collected. In unconstrained real-world scenario, the full face recognition approach is difficult because in many cases, the subject may have expression that distorts the face, or with closed eyes, or with facial occlusions as illustrated in Figure 1, 2, and 3. When the enrolled gallery image is frontal and constrained, it is very difficult to correctly match

Figure 1. Example of a masked suspect in ATM armed robbery (a), Taliban militant wearing a mask (b) and a bank robber wearing a mask (c) where only their periocular regions are visible for biometric identification.

to the highly unconstrained ones. In this work, we propose to use only the periocular region for face recognition because it has been proven by our evaluation that the periocular region has seen a great improvement over the full face for recognition across all kinds of unconstrained scenarios, *i.e.*, neutral expression, smiling expression, closed eyes, and facial occlusions. We collect the **Compass** database using the PTZ camera that reflects aforementioned unconstrained scenarios in the real world to a great extent. Using only the periocular region (with less than $2/5$ the size of the full face), we are able to obtain 60.7% verification rate (VR) at 0.1% false accept rate (FAR), a 16.9% improvement over the full face performance on images with all facial manners in the **Compass** database.

We have developed our work in the following fashion: Section 2 describes the hardware architecture of the PTZ camera system. Section 3 gives a step-by-step detail of how the periocular biometric is acquired. In Section 4, we discuss the building of a subspace representation and how the periocular recognition task is accomplished. Sections 5 detail our experimental setup for the verification experiments and report results from the experiments. Finally we conclude our work in Section 6.

2. System Hardware Architecture

In this section, we detail the hardware architecture of our proposed system.

Figure 2. Examples of crowds parading (a) and gathering (b) at Washington D.C. for President Obama's Inauguration Ceremony and crowds in subway station in Tokyo, Japan (c). There are many occluded faces in the crowd where only their periocular regions are visible for biometric identification.

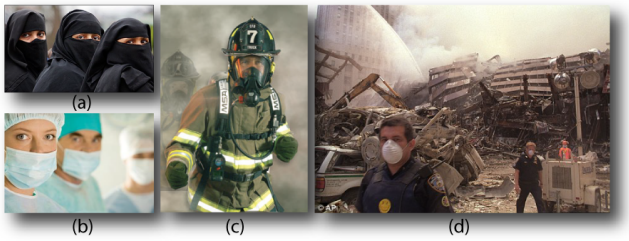


Figure 3. Examples of Muslim women wearing veil (a), doctors and nurses wearing clinical mask (b), fire fighter wearing air mask (c), and police officer wearing dust mask (d). People from certain religious groups and occupations occlude faces where only their periocular regions are visible for biometric identification.

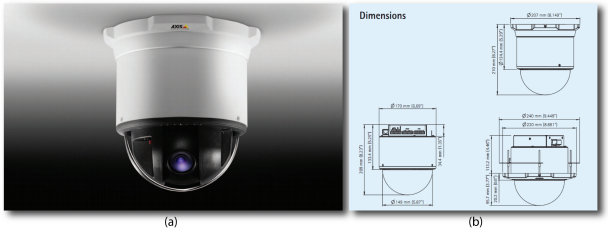


Figure 4. (a) The AXIS 233D Network Dome PTZ Camera used in this work and (b) its dimensions [1]

2.1. Pan-Tilt-Zoom Camera

The acquisition device used in the proposed system is the AXIS 233D Network Dome Pan-Tilt-Zoom (PTZ) camera (see Figure 4) manufactured by AXIS Communications. Details about the camera are available on the specification sheet [1]. The interesting specifications needed for our application include the following: the camera is built around a 1/4-inch ExView HAD progressive scan CCD. It is capable of $35\times$ optical zoom with a pan capability of 360° and a tilt capability of 180°. Both pan and tilt motions have adjustable speeds ranging from $0.05^\circ/\text{second}$ to $450^\circ/\text{second}$.

2.2. Acquisition Range

In order to get the periocular images that are suitable for recognition purposes, the inter-pupillary resolution must be at least 50 pixels. We know that a typical human inter-pupillary distance (IPD) is around $60\,\text{mm}$ [5]. The AXIS 233D network dome camera has a minimum focal length of $3.4\,\text{mm}$ and a maximum focal length of $119\,\text{mm}$. The AXIS 233D uses an ExView HAD sensor, which has a pixel size of $6.45\,\mu\text{m}$. If the periocular image needs to have an inter-pupillary resolution of at least 50 pixels, then the size of the eye-to-eye image on the sensor is given by:

$$v = 50 \times 6.45\,\mu\text{m} = 0.3225\,\text{mm}. \tag{1}$$

If we consider an average person, the IPD $u = 60\,\text{mm}$ [13], the magnification is given by:

$$M = \frac{\text{image size}(v)}{\text{object size}(u)} = \frac{0.3225}{60} = 0.0054. \tag{2}$$

The magnification of a lens system with effective focal length f , for an object at a distance of D in front of the lens is given by the standard relation:

$$M = \frac{f}{D - f}. \tag{3}$$

The far-end face images are captured using focal length of $119\,\text{mm}$, that is, $f = 119\,\text{mm}$. So,

$$M = 0.0054 = \frac{119}{D - 119} \Rightarrow D = 22.156\,\text{m}. \tag{4}$$

Hence, when the system is setup, the maximum allowable subject stand-off distance is approximately 22 meters if the required inter-pupillary resolution is at least 60 pixels.

3. Periocular Biometric Acquisition

This section outlines the processing steps and various algorithms used in this system.

3.1. Face Detection

The face detection is accomplished by the method proposed by Viola and Jones [17] in each camera frame. This method has been shown to provide good accuracy at high frame rate. There are three phases in this algorithm: (1) feature extraction, (2) classification using AdaBoost, and (3) multiscale detection.

Here we briefly introduce the three phases. Feature extraction for face detection utilizes Haar wavelet-like rectangular features of different sizes. AdaBoost is then adopted to build a "strong" classifier by combining the "weak" classifiers from each feature. AdaBoost takes a weighted vote of the decisions made by "weak" classifiers and outputs the final decision indicating whether face is presence or not. The way AdaBoost combines feature is by studying each feature at a time and finding which feature best classify the data. The samples correctly classified are given lower weights while those wrongly classified are given higher weights, and

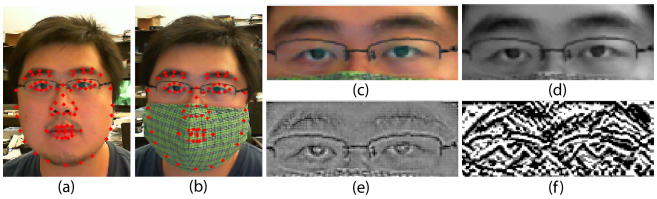


Figure 5. (a) 79 keypoints localized using ASM, (b) ASM still works on occluded face, (c) normalized periocular region, (d) gray-scale, (e) illumination preprocessing, and (f) WHT-LBP feature.

this updating process is iterated until convergence. The error is given by the sum of the weights of misclassified samples at any given iteration. The α_t shows how confident the t^{th} "weak" classifier is as:

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t}, \tag{5}$$

where ϵ_t is the classification error when using the t^{th} "weak" classifier. If $D_t(i)$ is the weight updated to the i^{th} sample after the previous iteration, then ϵ_t is given by:

$$\epsilon_t = \sum_{i=1}^{n} D_t(i) I(y_i \neq h_t(x_i)), \tag{6}$$

where $I()$ is the indicator function that satisfies:

$$I(y_i \neq h_t(x_i)) = \begin{cases} 1, & \text{if } y_i \neq h_t(x_i) \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where y_i is the true value of the i^{th} sample while $h_t(x_i)$ is the predicted value of this sample by the "weak" classifier h_t. The aforementioned phases are done in each scale in order to pick up faces of different sizes. This will allow AdaBoost to detect multiple faces at a given camera frame. Such characteristics is highly desirable for our system for it covers such a big capture volume.

This algorithm will box each detected face in a rectangle and send it to the following facial landmarking step.

3.2. Facial Keypoints Localizing

In this system, we utilize active shape model (ASM) to accurately localize the facial key points as shown in Figure 5 (a), specifically, we adopt a robust modified ASM developed by [14] that gives 79 facial keypoints. In order to construct a statistical facial model, we take a subset from the MBGC 2008 database [3] and manually annotate these face images with landmarks to formulate the training set for the ASM. The following are the training stages of the ASM.

1. For each training image, the coordinates of all 79 facial key points are stored as a shape vector $\mathbf{\Phi} = [x_1, y_1, x_2, y_2, ..., x_N, y_N]$, where x_i and y_i are the coordinates of the i^{th} keypoint and N is the number of key points used.

2. Generalized Procustes analysis [6] is utilized to align all the shape vectors $\{\mathbf{\Phi_j}\}$ from the entire training set.

3. Principal component analysis is then applied on these shape vectors $\{\mathbf{\Phi_j}\}$. We keep the top 97% eigenvectors, and also the mean shape $\bar{\mathbf{\Phi}}$ which will be used as the initial fit for an unseen face image.

4. Profiling step is necessary to build a subspace of variations of these pixel intensities across all 79 keypoints for all training images. There are two ways to profile the individual keypoint so that we can generate statistical models of the pixel intensity around each keypoint. A 1D profile of a keypoint is constructed by sampling the intensities of pixels lying orthogonal to the shape boundary at that keypoint. Each line is 17 pixels in length as shown in Figure 6 (a). Then the normalized gradient of these profile lines are stored as a vector. The mean of such vectors for each keypoint across all training images is called the mean profile vector $\bar{\mathbf{p}}$, and $\mathbf{\Sigma_p}$ is the covariance matrix of all profile vectors. Both $\bar{\mathbf{p}}$ and $\mathbf{\Sigma_p}$ are computed for each keypoint at four scale levels of the image pyramid. Similarly, 2D profile can be computed by sampling the gradient of pixel intensities in a square region around each keypoint as shown in Figure 6 (b).

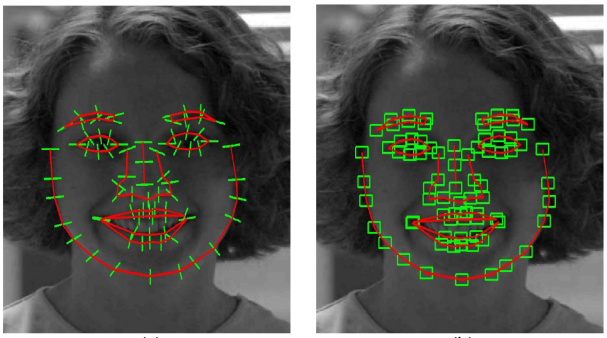
The following are the fitting stages of the ASM on unseen faces.

1. The mean shape $\bar{\mathbf{\Phi}}$ from the training stage is scaled, rotated and translated to best fit the unseen face as the initial shape model. It will then be deformed to achieve the final shape model associated with the unseen input face.

2. Profiling is performed in the same manner as in the training stage where multilevel profiles are constructed around each keypoint from coarse to fine as shown in Figure 7. At coarse level, larger step size is adopted to adjust the keypoint location while at fine level, smaller step size is adopted. The best location for a keypoint is determined by comparing profiles of neighboring patches adjacent to the candidate points [14]. Whichever candidate point has the most similar profile to the mean profile is considered as the new location for the keypoint. Here we adopt the Mahalanobis distance D as the similarity measurement:

$$D = (\mathbf{p} - \bar{\mathbf{p}}) \mathbf{\Sigma_p^{-1}} (\mathbf{p} - \bar{\mathbf{p}}). \tag{8}$$

The keypoint location updating process is repeated until the best location for each keypoint is reached.



(a) (b)

Figure 6. (a) 1D profile and (b) 2D profile illustrations. [14]

Figure 7. Multilevel pyramid to update keypoints location. [14]

When a newer shape is obtained, we can represent it by the vector $\mathbf{\Phi}_L$ which can be further represented by:

$$\mathbf{\Phi}_L = \bar{\mathbf{\Phi}} + \mathbf{V}\mathbf{c}, \tag{9}$$

where $\bar{\mathbf{\Phi}}$ is the mean shape estimated during the training stage, $\mathbf{V}$ is the eigenvector matrix from the ASM shapes determined during the training phase, and $\mathbf{c}$ is a vector of projection coefficients that needs to be calculated. This stage is necessary to make sure that the obtained shape is a legal shape modeled by the PCA face subspace [14]. This is done by iteratively minimizing the mean squared error $||\mathbf{\Phi}_L - T(\bar{\mathbf{\Phi}} + \mathbf{V}\mathbf{c})||^2$ during the testing phase. T is a similarity transform that minimizes the distance between $\mathbf{\Phi}_L$ and the shape given by $\bar{\mathbf{\Phi}} + \mathbf{V}\mathbf{c}$.

The keypoints are shifted at a particular pyramid level until no significant change in position is observed between two successive iterations. Following this, the keypoints are scaled and used as the initial positions for the next level of the pyramid. The process is terminated when there is convergence at the finest pyramid level. Thus, we get the final keypoint locations. The coarse-to-fine pyramid process is shown in Figure 7.

3.3. Periocular Region Normalization

Once the ASM model has been fit for the new face, the eye position can be obtained from the keypoints in the periocular region. In order to localize the periocular region

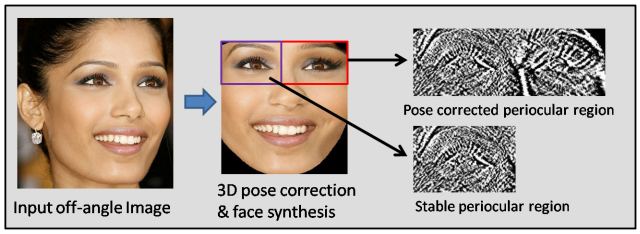


Figure 8. GEM method is applied to correct the off-angle pose for better periocular matching performance.

on images without pose correction, we utilize the 79 facial landmark points that are given by ASM to locate the eyes, and then rotation and eye coordinate normalization are performed to horizontally align left and right eyes with fixed eye coordinates for every image. On images with pose correction, since the faces are already well aligned, simple crop is sufficient. After the periocular region normalization, the entire strip containing both eyes is cropped in size of 50×128. Figure 5 (c) shows an example of the periocular region normalization on images without pose correction.

3.4. Pose Correction

When it comes to uncooperative or non-cooperative subjects, the image capture is non-ideal for most of the time, *i.e.*, the captured face image is off-angle. We utilize the 3D generic elastic models (GEMs) [12] to correct the off-angle pose for more robust periocular region recognition on uncooperative and non-cooperative subjects.

From any facial image captured by our system, we generate a 3D face for novel 2D pose synthesis. The depth of each 2D image is approximated from the canonical depth model while preserving the forensic information, and thus can correct the pose for matching. Examples of novel 2D pose correction using the GEMs are shown in Figure 8.

3.5. Illumination Preprocessing

To make this system even more robust, we here apply an illumination preprocessing step so that this system can match facial images taken under various illumination conditions. Illumination is the most significant factor affecting face appearance besides pose variation. The anisotropic diffusion model [7] has demonstrated excellent performance in challenging illumination conditions. Unfortunately, this algorithm is computationally demanding. To speed up the process, we follow a parallelized implementation of the anisotropic diffusion image preprocessing algorithm running on GPUs programmed with nVidia's CUDA framework and results are shown in Figure 5 (e).

3.6. Feature Extraction

In addition to directly applying LBP [10] on raw pixel for feature extraction, it's intuitively reasonable to post-apply

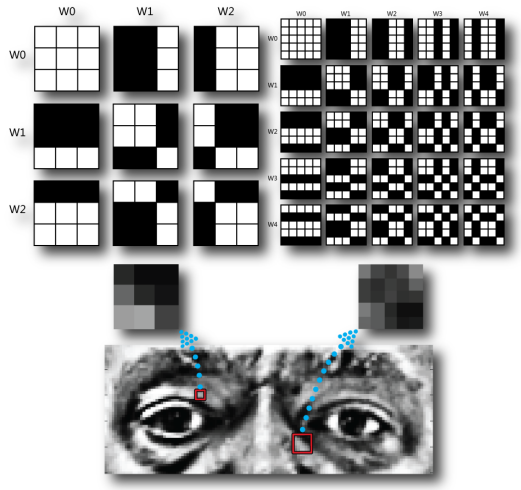


Figure 9. Expand each 5×5 kernels (right) using 25 Walsh basis images. The coefficients are treated as features assigned to the center pixel.

LBP upon other feature extraction techniques.

We introduce Walsh-Hadamard transform encoded local binary patterns (WHT-LBP) which is to apply LBP to encode the Walsh-Hadamard transform coefficients. In order to speed up the transform, we use Walsh masks as convolution filters to approximate the Walsh-Hadamard transform.

We utilize convolution filters such as Walsh masks [4, 11] to capture local image characteristics. The Walsh functions could be used to construct a complete and orthonormal basis in terms of which any matrix of a certain size may be expanded. All the basis of different sizes can be obtained by the Walsh function:

$$W_{2j+q}(t) = (-1)^{\lfloor \frac{2}{j} \rfloor + q}[W_j(2t) + (-1)^{j+q} W_j(2t-1)], \tag{10}$$

where $\lfloor \frac{2}{j} \rfloor$ means the integer part of $j/2$, q is either 0 or 1.

In the case of 5×5 kernel size, all possible combinations of Walsh vectors can be used to produce 25 2D basis images as shown in Figure 9.

Now, we can represent an image patch using Walsh elementary vectors of even size:

$$g = WfW^T, \tag{11}$$

where the transformation matrix W, the input image patch f and the output coefficients g are all of the same size, $N \times N$ (N is even). An even size transformation matrix constructed from Walsh functions is orthogonal thus its inverse is its transpose ($W^{-1} = W^T$).

In our experiment, we use image kernel of odd size such as 3×3, 5×5 and 7×7. As mentioned before, odd-sized Walsh vectors yield an odd-sized Walsh transformation matrix. Such matrix is no longer orthogonal ($W^{-1} \neq W^T$). In order to invert Equation 11, we make use of the inverse of W. Then we shall have [11] $W^{-1}g(W^T)^{-1} = f$.

So, we may use the inverse of matrix W to process the image. Then we have $g = (W^{-1})^T f W^{-1}$.

Once we have the Walsh coefficients, we post-apply LBP upon those coefficients and create WHT-LBP. The WHT-LBP feature can dramatically improve the verification rate than LBP especially on the periocular region [8]. Examples of the WHT-LBP feature on the periocular image are shown in Figure 5 (f).

4. Periocular Biometric Recognition

In this section, we detail a subspace representation technique: the kernel class-dependence feature analysis (KCFA) [16] for the periocular biometric matching.

4.1. Correlation Filter Theory

Correlation filter technique [15] has been widely used for frequency domain image processing. Mahalonobis et al. [9] developed minimum average correlation energy (MACE) filter which is designed to minimize the average correlation energy resulting from the training images while exhibiting strong peak at the location of a trained object. The closed vector form of the resulting MACE filter is:

$$\mathbf{h} = \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^{+}\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{u}, \tag{12}$$

where $\mathbf{X}$ is $d^2 \times N$ matrix containing 1D column vectors that are converted from 2D FFT arrays from N training images. Here, N is the number of training images and d^2 is the number of pixels in each image. $\mathbf{D}$ is $d^2 \times d^2$ diagonal matrix containing average power spectrum of the training images on its diagonal. Column vector $\mathbf{u}$ contains the N pre-specified correlation peak values at the origin. The + symbol denotes the complex conjugate transpose.

4.2. Class-Dependence Feature Analysis

The class-dependence feature analysis (CFA) method has shown its ability to perform robust face recognition in well represented training data. The CFA method [16] uses a set of MACE filters [9] to extract features from the individuals in the training set. MACE filters are generated for every subject in the generic training set therefore number of class matches the dimensionality of the feature space. The preset correlation peaks ensure that the filter finds no correlation between subjects that belongs to different classes. The correlation of a test image $\mathbf{y}$, with N MACE filters can be expressed as:

$$\mathbf{c} = \mathbf{H}^{\top}\mathbf{y} = [\mathbf{h}_{\text{MACE}-1}\mathbf{h}_{\text{MACE}-2}\cdots\mathbf{h}_{\text{MACE}-N}]^{\top}\mathbf{y}, \tag{13}$$

where $\mathbf{h}_{\text{MACE}-i}$ is a MACE filter that has preset value of the correlation peak to give a small correlation output that

is close to zero for all classes except for class-*i*. Each input image is then projected onto those correlation feature vector $\mathbf{c}$ where N is the number of training classes.

In our method, we use only a small portion (444/12776) of the generic training set in FRGC ver2.0 database [2]. One MACE filter is designed for each of the 222 subjects in FRGC generic training set. The design of these 222 filters utilizes only 2 images per subject in one-class-against-the-rest fashion. When 222 filters are designed, they are used as projection bases to produce the output feature vector $\mathbf{c}$ for any test image $\mathbf{y}$ in Equation 13.

4.3. Kernel CFA

The KCFA [16] is designed to overcome the low performance of the linear subspace approach due to the nonlinear distortions in human face appearance variations. The features in the higher dimensional space are obtained using inner products in the linear space, without actually forming the higher dimensional feature mappings. This kernel trick improves efficiency and keeps the computation tractable even with the high dimensionality. The mapping function can be denoted as: $\mathbf{\Phi} : R^N \rightarrow F$. Kernel functions are defined by:

$$K(\mathbf{x}, \mathbf{y}) = \langle \mathbf{\Phi}(\mathbf{x}), \mathbf{\Phi}(\mathbf{y}) \rangle, \tag{14}$$

which can be used without having to form the mapping $\mathbf{\Phi}(\mathbf{x})$ as long as kernels form an inner product and satisfy Mercer's theorem [18]. Examples of kernel functions are:

1. Polynomial: $K(\mathbf{x}, \mathbf{y}) = \big(\langle \mathbf{x}, \mathbf{y} \rangle + 1\big)^p$.
2. RBF: $K(\mathbf{x}, \mathbf{y}) = \exp\big(-(\mathbf{x} - \mathbf{y})^2 / 2\sigma^2\big)$.
3. Sigmoidal: $K(\mathbf{x}, \mathbf{y}) = \tanh\big(\kappa \langle \mathbf{x}, \mathbf{y} \rangle - \delta\big)$.

Pre-filtering process is performed before the derivation of the closed form solution to the kernel MACE filter. The correlation output of a filter $\mathbf{h}$ and a test image $\mathbf{y}$ can be expressed as:

$$\begin{aligned} \mathbf{y}^+\mathbf{h} &= \mathbf{y}^+[\mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^+\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{u}] \\ &= (\mathbf{D}^{-\frac{1}{2}}\mathbf{y})^+(\mathbf{D}^{-\frac{1}{2}}\mathbf{X})\big((\mathbf{D}^{-\frac{1}{2}}\mathbf{X})^+ \cdot \mathbf{D}^{-\frac{1}{2}}\mathbf{X}\big)^{-1}\mathbf{u} \\ &= \big((\mathbf{y}')^+\mathbf{X}'\big)\big((\mathbf{X}')^+\mathbf{X}'\big)^{-1}\mathbf{u}, \end{aligned} \tag{15}$$

where $\mathbf{X}' = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}$ denotes pre-whitened $\mathbf{X}$. Now we can apply the kernel trick to yield the kernel correlation filter:

$$\begin{aligned} \mathbf{\Phi}(\mathbf{y}) \cdot \mathbf{\Phi}(\mathbf{h}) &= \big(\mathbf{\Phi}(\mathbf{y}) \cdot \mathbf{\Phi}(\mathbf{X})\big)\big(\mathbf{\Phi}(\mathbf{X}) \cdot \mathbf{\Phi}(\mathbf{X})\big)^{-1}\mathbf{u} \\ &= K(\mathbf{y}, \mathbf{x}_i)K(\mathbf{x}_i, \mathbf{x}_j)^{-1}\mathbf{u}. \end{aligned} \tag{16}$$

The KCFA method completes after extracting 222-dimensional kernel feature vectors using the kernel correlation filters in the same CFA framework. We input WHT-LBP feature into KCFA subspace modeling instead of raw pixel intensity to yield better recognition performance.

Table 1. **Compass** Database Specifics

Ethnicity	**Male**	**Female**	**Total**
East Asian	4	4	8
South Asian	8	7	15
Caucasian	6	4	10
African	4	3	7
Total	22	18	**40**

5. Experimental Results

In this section, we first detail the experimental setup on our own database and then report the corresponding results.

5.1. Setup

There has not been any public database that serves this study on the periocular biometric for uncooperative and non-cooperative subjects using the PTZ camera. So we decide to create our own database and compare the performance on the periocular region against the full face. In order to better capture the real-life uncooperative and non-cooperative scenario, subjects to be put into our database are asked to present the following four facial manners: **N** - neutral expression, **S** - smiling expression, **E** - eyes closed, and **W** - with facial occlusions (subjects can cover nose, mouth, or cheek), at two distances: $10\,\mathrm{m}$ and $20\,\mathrm{m}$. With the four initials shown above **N**, **S**, **E**, **W**, we decide to name our database **Compass**. Examples of database images are demonstrated in Figure 10.

Totally, there are 40 subjects captured in our database, each has 10 images per facial manner at each distance. So there are 80 images per subject. This leads to 3,200 images in the entire database. Table 1 shows the specifics of the **Compass** database. The table shows that we strike a balance among different ethnicity groups and between genders. We intentionally make the database in this fashion so that it can better represent the real world scenario.

Each periocular image is acquired, preprocessed, normalized, and extracted with features following the steps described in Section 3 and then project onto the 222 filters designed by utilizing 444 training images in the FRGC ver2.0 database as described in Section 4, so each periocular image is converted into a 222-dimension feature vector. Similarly, we follow the same steps to process the full face image whose performance will be used as the benchmark for our system that utilizes the periocular biometric only.

For all the experiments we carried out, normalized cosine distance (NCD) measurement is adopted to compute similarity matrix:

$$d(\mathbf{x}, \mathbf{y}) = \frac{-\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}. \tag{17}$$

Compared to other commonly used distance measurement such as L1-norm, L2-norm, the Mahalanobis distance, NCD

Figure 10. Examples from Compass database. Images are taken at 10 meters (top row) and 20 meters (bottom row), with 4 different facial manners.

exhibits the best result.

The result is a similarity matrix with the size of $N \times N$ whose entry SimM_{ij} is the NCD between the feature vector of probe image i and target image j. The performance is analyzed using verification rate (VR) at 0.1% false accept rate (FAR) and the receiver operating characteristic (ROC) curves.

5.2. Results

In this subsection, we report the experimental results. We perform all the verification experiments on both the periocular region with the size of 50×128 and the full face with the size of 128×128 on each partition of the **Compass** database at 10 meters and 20 meters.

Table 2 shows the verification rate at 0.1% false accept rate. The pair of letter/number in black font like **Target**, **Probe** means that the first parameter is treated as target image while the second parameter probe image. Specifically, **All**, **All** in the last column means images with all 4 facial manners are matched against each other $(3,200 \times 3,200)$ This can best characterize the real life scenario where face images from uncooperative and non-cooperative subjects are unconstrained.

Figure 11 shows the ROC curves for experiments on all 4 facial manners, *i.e.*, the last column in Table 2. The results clearly show that when it comes to real world scenario where face images are highly unconstrained with expression, closed eyes, and facial occlusions from uncooperative and non-cooperative subjects, recognition using only the periocular region performs much better than the full face.

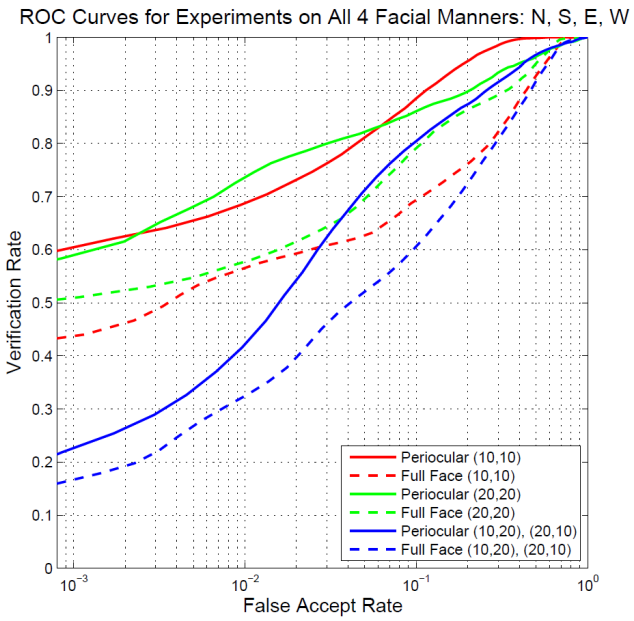


Figure 11. ROC curves for matching across all facial manners.

6. Conclusion

In this work, we have proposed an acquisition and recognition system based only on the periocular biometric using commercial off-the-shelf pan-tilt-zoom camera to tackle the difficulty that the full face recognition approach has encountered in highly unconstrained real-world scenario, especially for capturing and recognizing uncooperative and non-cooperative subjects with expression, closed eyes, and facial occlusions. We evaluate our algorithm on the periocular region and compare that to the performance of the full face on the **Compass** database we have collected. The results have shown that the periocular region, when tackling unconstrained matching, is a much better choice than the full face for face recognition even with less than $2/5$ the size of the full face. To be more specific, on our **Compass** database, the periocular matching across all facial manners, *i.e.*, neutral expression, smiling expression, closed eyes, and facial occlusions, is able to achieve 60.7% verification rate at 0.1% false accept rate, a 16.9% performance boost over the full face. Based on our verification experiments, we

Table 2. Verification Rate at 0.1% False Accept Rate on **Compass** Database

Target, Probe	**N, N**	**N, S**	**N, E**	**N, W**	**S, S**	**S, E**	**S, W**	**E, E**	**E, W**	**W, W**	**All, All**
Periocular **10, 10**	0.939	0.849	0.283	0.204	0.994	0.239	0.369	0.869	0.150	0.848	**0.607**
Periocular **20, 20**	0.917	0.734	0.121	0.191	0.946	0.250	0.216	0.890	0.093	0.739	**0.589**
Periocular **10, 20**	0.304	0.344	0.048	0.242	0.441	0.044	0.309	0.299	0.149	0.408	**0.227**
Periocular **20, 10**	0.304	0.328	0.184	0.068	0.441	0.147	0.089	0.299	0.044	0.408	**0.227**
Full face **10, 10**	0.944	0.582	0.684	0.087	0.995	0.576	0.126	0.968	0.203	0.889	0.438
Full face **20, 20**	0.943	0.819	0.714	0.216	0.981	0.547	0.232	0.933	0.075	0.969	0.510
Full face **10, 20**	0.341	0.199	0.153	0.196	0.207	0.163	0.074	0.276	0.174	0.509	0.170
Full face **20, 10**	0.341	0.301	0.332	0.144	0.207	0.250	0.101	0.276	0.097	0.509	0.170

can safely draw the conclusion that the periocular biometric outperforms the full face under unconstrained scenarios for uncooperative and non-cooperative subjects.

In future, we will keep on collecting more samples to build a much larger **Compass** to better reflect the unconstrained real-world scenario.

References

[1] Axis 233d camera datasheet. http://www.axis.com/products/cam_233d/index.htm. 202

[2] Face recognition grand challenge (frgc). http://www.nist.gov/itl/iad/ig/frgc.cfm. 206

[3] Multiple biometric grand challenge (mbgc). http://face.nist.gov/mbgc/. 203

[4] K. Beauchamp. Applications of walsh and related functions. In *Academic Press*, 1984. 205

[5] J. Forrester, A. Dick, P. Mcmenamin, and W. Lee. *The Eye: Basic Sciences in Practice*. W. B. Saunder, London, UK, 2001. 202

[6] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975. 203

[7] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *4th Int'l Conf. on Audio- and Video-Based Biometric Person Authentication*, pages 10–18, 2003. 204

[8] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, and C. Suen. Investigating age invariant face recognition based on periocular biometrics. In *International Joint Conference on Biometrics*, oct 2011. 205

[9] A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent. Minimum average correlation energy filters. *Appl. Opt.*, 26(17):3633–3640, Sep 1987. 205

[10] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. In *Pattern Recognition*, volume 29, pages 51–59, 1996. 204

[11] M. Petrou and P. Sevilla. *Dealing with Texture*. Wiley, 2006. 205

[12] U. Prabhu, J. Heo, and M. Savvides. Unconstrained pose-invariant face recognition using 3d generic elastic models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1952–1961, oct. 2011. 204

[13] S. Ray. *Applied Photographic Optics, 3ed*. Focal Press, 2002. 202

[14] K. Seshadri and M. Savvides. Robust modified active shape model for automatic facial landmark annotation of frontal faces. In *Proceedings of the IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems (BTAS'09)*, 2009. 203, 204

[15] B. V. K. Vijaya Kumar, A. Mahalanobis, and R. D. Juday. *Correlation Pattern Recognition*. Cambridge University Press, UK, 2005. 205

[16] B. V. K. Vijaya Kumar, M. Savvides, and C. Xie. Correlation pattern recognition for face recognition. *Proc. of the IEEE*, 94(11):1963–1976, nov 2006. 205, 206

[17] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1511–1518, 2001. 202

[18] M.-H. Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *Automatic Face and Gesture Recognition. Proc. 5th IEEE Int'l Conf. on*, pages 215–220, may 2002. 206