RESEARCH **Open Access**

LBP-based periocular recognition on challenging face datasets

Gayathri Mahalingam[*] and Karl Ricanek Jr

Abstract

This work develops a novel face-based matcher composed of a multi-resolution hierarchy of patch-based feature descriptors for periocular recognition - recognition based on the soft tissue surrounding the eye orbit. The novel patch-based framework for periocular recognition is compared against other feature descriptors and a commercial full-face recognition system against a set of four uniquely challenging face corpora. The framework, hierarchical three-patch local binary pattern, is compared against the three-patch local binary pattern and the uniform local binary pattern on the soft tissue area around the eye orbit. Each challenge set was chosen for its particular non-ideal face representations that may be summarized as matching against pose, illumination, expression, aging, and occlusions. The MORPH corpora consists of two mug shot datasets labeled Album 1 and Album 2. The Album 1 corpus is the more challenging of the two due to its incorporation of print photographs (legacy) captured with a variety of cameras from the late 1960s to 1990s. The second challenge dataset is the FRGC still image set. Corpus three, Georgia Tech face database, is a small corpus but one that contains faces under pose, illumination, expression, and eye region occlusions. The final challenge dataset chosen is the Notre Dame Twins database, which is comprised of 100 sets of identical twins and 1 set of triplets. The proposed framework reports top periocular performance against each dataset, as measured by rank-1 accuracy: (1) MORPH Album 1, 33.2%; (2) FRGC, 97.51%; (3) Georgia Tech, 92.4%; and (4) Notre Dame Twins, 98.03%. Furthermore, this work shows that the proposed periocular matcher (using only a small section of the face, about the eyes) compares favorably to a commercial full-face matcher.

1 Introduction

The field of biometrics has made significant accomplishments over the last 20 years. Biometric systems are now deployed in dozens of countries for a host of purposes from national identification to access, to amusement parks, to automatic log in for computing devices. As the technology matures, users demand better performance against non-ideal (poor) biometric signals, e.g., border crossing systems should be able to capture the biometric signal of the iris or face while patrons are moving or computers should be able to authenticate patron credentials 10 years or more after enrollment without the requirement of template updating. Deployers as well as end users of biometric systems demand more flexibility in acquiring the biometric signal and better performance against matching to biometric templates that differ due to pose, illumination, expression, and aging.

Non-ideal biometrics, which is also known as unconstrained biometrics, are systems that do not force (constrain) the user to submit their biometric signal (face, iris, fingerprint, etc.) in a purposed manner. Furthermore, they are systems that can perform robust matching against templates that have been acquired under non-ideal or bad conditions. Non-ideal face recognition is recognition systems that are capable of matching well against probe images that may exhibit poor image quality, low image resolution, poor lighting, occlusions and disguises, heavy pose variation, and/or moderate to severe expression or face contortions. Non-ideal face must also contend with aging and the challenges of matching under aging as well as dealing with the case of matching in the presence of extremely similar faces, i.e., discriminating between identical twins.

Periocular-based recognition has gained increasing attention from biometric researchers recently. Park et al. [1] studied the use of the periocular region as a useful

*Correspondence: mahalingamg@uncw.edu
ISIS Institute-Face Aging Group, University of North Carolina Wilmington, Wilmington, NC, USA

Table 1 Rank-1 match performance on challenging datasets

Datasets (rank-1 accuracy)	Face recognition literature (%)	COTS (%)	Proposed (%)
Georgia Tech DB [12]	96.60 [16]	99.73	92.42
Twins Face DB [13]	96.24[a] [17]	98.04	98.03
MORPH Album1 [15]	50.4[b] [18]	51.60	33.20
FRGC [14]	#	99.65	97.51

The table lists the rank-1 accuracies obtained for Georgia Tech face database, Twins face database, MORPH Album 1, and FRGC database against the literature, COTS matcher used in this work, and the proposed algorithm. The number sign indicates that [14] performance measures against the FRGC are based on verification rate of 98 at 0.1% FAR. [a]The results are from both identical and fraternal twins for [17], while our approach used only the identical twins. [b]Average of rank-1 accuracies of all age groups with an age gap of 0 to 5 years between the probe and gallery.

biometric when iris recognition fails. The authors proposed a matching scheme with three descriptors: gradient orientation, local binary pattern (LBP), and SIFT. Their experimental comparison of periocular-based recognition with that of face recognition under occlusion showed superior performance of the periocular recognition system. Similarly, [2] and [3] illustrated the effectiveness of periocular-based features for recognition using images focused on capturing the iris. Both the periocular skin texture [4] and its appearance cues [5,6] were used for recognition. Padole and Proenca [7] studied the performance of a periocular-based recognition system under the influence of scale, pose, occlusion, etc. and concluded that the performance of the recognition system degrades with the presence of such covariates. Xu et al. [8] proposed an age-invariant recognition system based on periocular features against a small dataset of longitudinal images.

Periocular features have also been used to identify other soft biometric cues such as gender [9], which showed the use of shape-based features of the eyebrow for biometric recognition and for gender classification purposes. Studies indicate the usefulness of such features for the task of verification by humans using near-infrared periocular images [10,11]. Although prior works have studied the performance of periocular-based

features under various scenarios, no work has focused on recognition performance in challenging real-world datasets that include images captured under extreme conditions, e.g., occlusions, poor lighting, being scanned from hard copy photographs, pose variations, aging, or twins.

In this work, we present a multi-scale, center-symmetric, patch-based LBP framework for recognition using four distinctive and challenging datasets. The proposed framework allows for effective description of the periocular features and matching them. The framework is evaluated on the Georgia Tech face database [12], Notre Dame (ND) Twins face database [13], FRGC [14], and the MORPH *Album 1* database [15]. These datasets include face images with variations in pose, illumination, expressions, eyewear, and some motion blur. The images of the Georgia Tech and the ND Twins database are digital photographs, while the images of MORPH Album 1 are scanned legacy photographs. MORPH Album 1 contained images of heavy occlusions across the face, poor (low) contrast and dynamic range, yellowing and cracking of source photographs, and many more challenges. Of the four datasets, MORPH Album 1 was the most difficult for all the algorithms tested. Our work analyzes the effectiveness of LBP-based feature descriptors on such datasets for
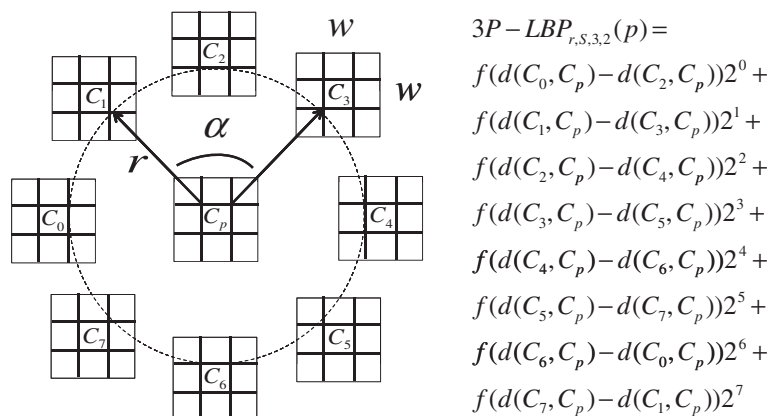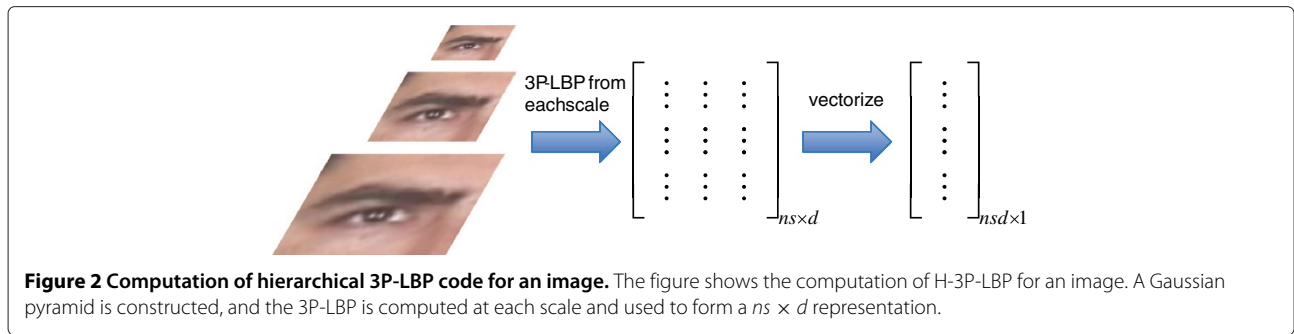

$$3P-LBP_{r,S,3,2}(p) =$$
$$f(d(C_0,C_p) - d(C_2,C_p))2^0 +$$
$$f(d(C_1,C_p) - d(C_3,C_p))2^1 +$$
$$f(d(C_2,C_p) - d(C_4,C_p))2^2 +$$
$$f(d(C_3,C_p) - d(C_5,C_p))2^3 +$$
$$f(d(C_4,C_p) - d(C_6,C_p))2^4 +$$
$$f(d(C_5,C_p) - d(C_7,C_p))2^5 +$$
$$f(d(C_6,C_p) - d(C_0,C_p))2^6 +$$
$$f(d(C_7,C_p) - d(C_1,C_p))2^7$$

Figure 1 Computation of three-patch LBP code for a pixel. The figure shows the computation of three-patch LBP code for a pixel.

Figure 2 Computation of hierarchical 3P-LBP code for an image. The figure shows the computation of H-3P-LBP for an image. A Gaussian pyramid is constructed, and the 3P-LBP is computed at each scale and used to form a $ns \times d$ representation.

periocular recognition. Our work differs from other earlier works in both the framework and a detailed analysis on a very difficult legacy dataset, MORPH Album 1. Our detailed analysis on the performance of both commercial and noncommercial recognition algorithms provide an insight on possible improvements to existing algorithms to better learn facial features in order to improve recognition under non-ideal conditions. Table 1 lists the rank-1 face recognition accuracies obtained in literature for the images from these datasets and the rank-1 periocular recognition accuracies obtained from the proposed framework on these datasets. The results indicate that the proposed framework provides a performance comparable to those of the commercial full-face recognition system used in this work.

The rest of the paper is organized as follows: Section 2 provides a detailed explanation of the proposed periocular recognition framework and the hierarchical three-patch local binary pattern (H-3P-LBP). Section 3 addresses the experiments conducted under covariates, including the experimental setup, the datasets used, the preprocessing steps, and the results. Section 4 provides the conclusions drawn and future work.
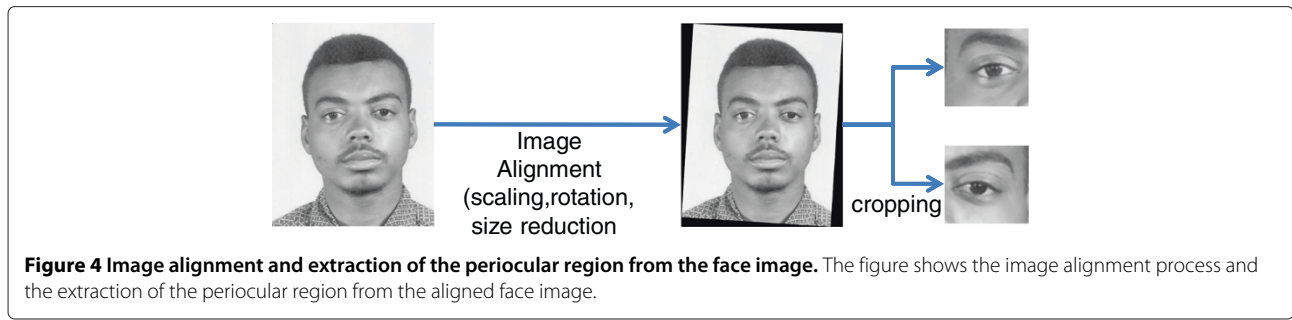
2 Periocular biometrics

The task of recognition (face or periocular) generally includes the following sequence of steps: preprocessing (image alignment, noise removal, illumination correction, etc.), feature extraction, and matching. For this study, we compute the periocular feature descriptors using uniform LBP [19], 3P-LBP [20], and its variant hierarchical three-patch local binary pattern (H-3P-LBP) (proposed). The uniform LBP follows a pixel-based approach with respect to computing the LBP code of a pixel and its neighboring sampling points, while the other two descriptors are patch-based approaches. Patch-based computation of texture patterns encodes the similarities between neighboring patches of pixels and thus captures information which is complementary to that of pixel-based descriptors. The patch-based textures treat colored regions, edges, lines, and complex textures in a unified way unlike pixel-based techniques.

2.1 Feature description using hierarchical 3P-LBP

The H-3P-LBP extends the 3P-LBP operator [20] by computing over different scales (multi-resolutions) of an image. The 3P-LBP of a pixel is computed by comparing the values of three patches to produce a single bit value in the code assigned to the pixel. 3P-LBP for each pixel is computed by considering a window of region centered on the pixel and considering m sampling points within a radius of r pixels. Unlike LBP, the 3P-LBP approach considers m patches around m neighboring pixels that are distributed uniformly around the center patch. The patch-based comparison in 3P-LBP is done by comparing the value of the center patch with a pair of patches that are α patches apart along the circle. The value of a single bit is set according to the similarity of the two patches with

Figure 3 Sample images from the datasets. The figure shows the sample images from Georgia Tech, Twins, and MORPH Album 1 datasets, respectively. The images show variations in pose, illumination, image artifacts, expressions, etc.

Figure 4 Image alignment and extraction of the periocular region from the face image. The figure shows the image alignment process and the extraction of the periocular region from the aligned face image.

the center patch. The resulting code has m bits per pixel. Figure 1 shows the computation of 3P-LBP code for a pixel.

The 3P-LBP is given by the following equation:

$$3\text{P-LBP}_{r,m,w,\alpha}(p) = \sum_{i}^{m} (f(d(C_i, C_p) - d(C_{i+\alpha \mod m}, C_p))) 2^i \quad (1)$$

where C_i and $C_{i+\alpha}$ are two patches along the ring and C_p is the central patch. The function $d(.,.)$ is any distance function between two patches (e.g., L_2 norm of their gray level differences), and f is defined as

$$f(x) = \begin{cases} 1 & \text{if } x \geq \tau \\ 0 & \text{if } x < \tau \end{cases} \quad (2)$$

where τ is set to a value slightly larger than zero in order to provide stability in uniform regions as indicated in [21].

For a given image I, a Gaussian pyramid with s levels is constructed to form the multi-scale representation of the image, with I being the finest scale in the Gaussian pyramid. The H-3P-LBP descriptor is computed by applying the 3P-LBP operator at each level of the image pyramid.

The final H-3P-LBP descriptor $H(I)$ is obtained by combining the 3P-LBP descriptors from each level into a final feature matrix. The hierarchical 3P-LBP $H(I)$ maps the image I into a $\mathbb{R}^{ns \times d}$ representation, where d is the length of the 3P-LBP code per pixel, and ns is the sum of the number of codes obtained from each image scale. In our experiments, we construct the image at three different scales. Figure 2 shows the computation of H-3P-LBP for an image.

The multi-scale 3P-LBP can be extracted either by varying the radii or by extracting 3P-LBP from different image scales. However, the first approach has its own shortcomings in the way the conventional 3P-LBP is applied to the image. The conventional approach typically extracts microstructures (edges, corners, spots, etc.) of the images, while the hierarchy allows for the extraction of both micro- and macro-structures [22], which are required for effective texture extraction and discrimination. The stability of 3P-LBP decreases with the increase in neighborhood radii due to minimal correlation of the sampling points with the center pixel. Also, the sparse sampling by 3P-LBP from a large neighborhood radii may not result in an adequate representation of the two-dimensional image signal.

Table 2 Rank-1 accuracies ND Twins effect of periocular image size and image scale

Pyramid levels (P), image size ($S \times S$)	Approach	Left (%)	Right (%)	α-Fusion (%)
$P = 4, S = 256$	H-3P-LBP	96.60	98.62	*98.11*
	LBP	52.83	56.23	66.79
$P = 4, S = 128$	H-3P-LBP	96.60	98.62	*98.11*
	LBP	52.83	56.23	67.30
$P = 3, S = 256$	H-3P-LBP	88.72	84.85	*90.49*
	LBP	55.56	51.01	65.32
$P = 3, S = 128$	H-3P-LBP	97.42	98.33	*98.03*
	LBP	53.18	52.73	70.15
$P = 3, S = 100$	H-3P-LBP	86.95	85.52	*89.8*
	LBP	56.31	51.94	64.90
$P = 1, S = 256$	3P-LBP	96.98	96.60	*98.11*
$P = 1, S = 128$	3P-LBP	96.98	96.60	*98.11*

The accuracies for LBP are achieved using the periocular image from the finest scale of the image pyramid. The 3P-LBP was achieved with $P = 1$ (the last two rows of the table). Values in italics are the best achieved results.

Table 3 Rank-1 accuracies for the left and right periocular regions

Datasets	Eye	H-3P-LBP (%)	3P-LBP (%)	LBP (%)
MORPH Album1	Left	*22.77*	20.62	2.22
	Right	*29.81*	26.77	3.11
Georgia Tech	Left	*86.87*	85.02	53.78
	Right	*90.74*	88.13	56.98
ND Twins	Left	*97.42*	96.36	53.18
	Right	*98.33*	98.18	52.73
FRGC	Left	*97.44*	97.23	82.99
	Right	96.97	*97.00*	82.66

The table indicates the rank-1 accuracies for both the left and right periocular regions for LBP, 3P-LBP, and H-3P-LBP descriptors. Values in italics are the best achieved results.

These observations are verified from the experimental results of various challenging datasets.

2.2 Match score generation

For this study, the Euclidean distance measure is used to formulate the match score between a pair of features. In addition, a score-level fusion is adapted to fuse the scores from the left and the right periocular region. The final score, α score, is used as the similarity measure between the probe and target set. The fusion of the scores for the left (S_l) and the right (S_r) periocular region is given by

$$S = \alpha \times S_l + (1 - \alpha) \times S_r \qquad (3)$$

where α denotes the weighting factor. The optimal value of α was determined off-line using a grid search method based upon randomly selected subset of the four datasets used for this work. The α value used in this work is 0.7. The match scores are fused using the weighted sum rule without any score normalization. Earlier research work [23] suggests that the recognition accuracy of the left periocular region (left from the observers' perspective) is significantly higher than that of the right periocular region. Although it has been shown that the left periocular region is more discriminative than the right, the reasoning behind this observation needs further investigation. The selected weighting factor is in accordance with this observation, providing more weight to the left periocular region.

3 Experiments

This section provides a detailed discussion on the datasets used for the study, the recognition experiments, and their results.

3.1 Database

The following databases were used in our experiments. These databases include face images of subjects taken under unconstrained conditions such as variations in pose, expression, and illumination; presence of glasses; facial hair; occlusions; etc. Also, these databases are publicly available and, hence, are more suitable for the research community to evaluate against and compare their results to.

3.1.1 Georgia Tech face database

The Georgia Tech face database (DB) [12] includes images of 50 subjects taken at multiple sessions. The database consists of 750 images with 15 images per subject. The images for each subject include the following variations: frontal pose, titled face with different facial expressions, illumination variations, and scale. The images are taken at a resolution of 640 × 480 pixels. Figure 3 shows sample images of a subject showing the abovementioned variations.

3.1.2 MORPH aging database

MORPH Album1 [15] consists of 1,690 scanned photographs of 515 individuals taken over an interval of time. The age of the images range from 15 to 68 years with the age gap between the first image and the last image ranging from 46 days to 29 years. The face images of Album 1 are frontal or near-frontal images under many types of illumination and eye region occlusions. This album of the MORPH dataset has been used for several years to evaluate the performance of recognition under aging [24,25]. Figure 3 shows sample images from the MORPH database illustrating the various challenges involved with the images.
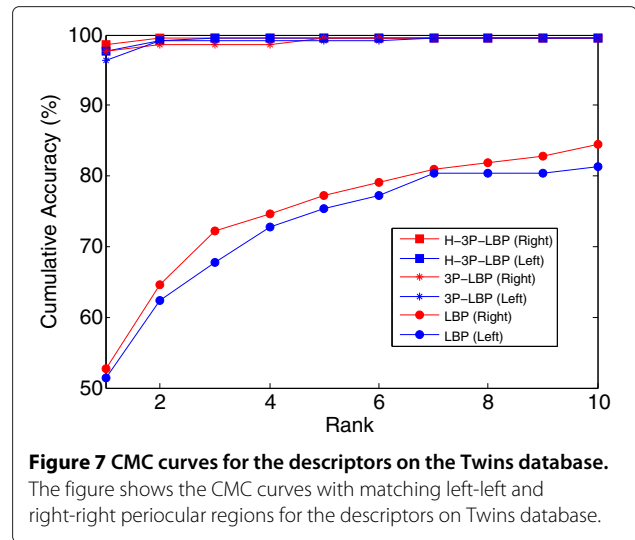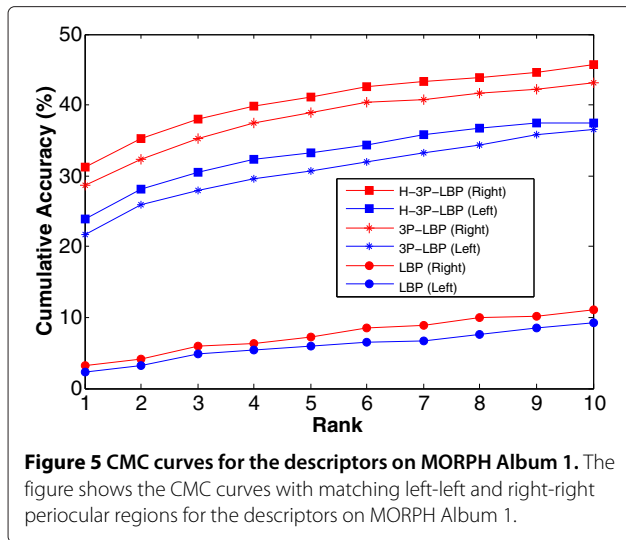
3.1.3 WVU/ND twins database

The Twins database [13] is comprised of multi-modal biometric information from pairs of identical and fraternal twins who attended the 2010 Twins Day Festival in Twinsburg, Ohio. The database consists of 6,863 2D color

Table 4 Rank-1 accuracies using fusion of scores from the left and right periocular regions

	MORPH Album 1 (%)	Georgia Tech (%)	Twins DB (%)	FRGC (%)
LBP	3.28	67.85	70.15	90.03
3P-LBP	30.74	91.41	*98.48*	97.44
H-3P-LBP	*33.20*	*92.42*	98.03	*97.51*

The table indicates the rank-1 accuracies obtained by fusion of scores from the left and right periocular regions for LBP and its variants. Values in italics are the best achieved results.

Figure 5 CMC curves for the descriptors on MORPH Album 1. The figure shows the CMC curves with matching left-left and right-right periocular regions for the descriptors on MORPH Album 1.

Figure 7 CMC curves for the descriptors on the Twins database. The figure shows the CMC curves with matching left-left and right-right periocular regions for the descriptors on Twins database.

face images from 240 subjects and were collected under varying lighting conditions (indoor/outdoor), expression (neutral/smile), and pose (frontal/non-frontal). Each image is of resolution 600×400 pixels. For our experiments, we used only the images of 100 pairs of identical twins and a triplet. (The identical twins/triplet images were used solely due to the very difficult nature of matching against them.) The images with neutral expression and with a frontal pose alone were included. Figure 3 shows sample images from the Twins face database.

3.1.4 FRGC database
The FRGC database includes around 16,000 images of 466 subjects collected at the University of Notre Dame during the academic years 2002 to 2003 and 2003 to 2004. The images for a subject session include four controlled

still images, two uncontrolled still images, and a three-dimensional image. The controlled images were taken under studio lighting conditions and two facial expressions (neutral and smile).

3.2 Image alignment
The periocular images can be aligned using certain key points such as eye centers, eye corners, eyelids, etc., which are some common components of the periocular region that can be identified fairly easily. The eye centers are good candidates as they can be identified with periocular images involving large pose variations, while the other key points suffer from occlusion due to pose changes. The motion of the iris and the eyelids are not significant in the periocular images used in this research. Therefore, we primarily used the eye centers for image alignment. The eye centers were detected using the commercial software
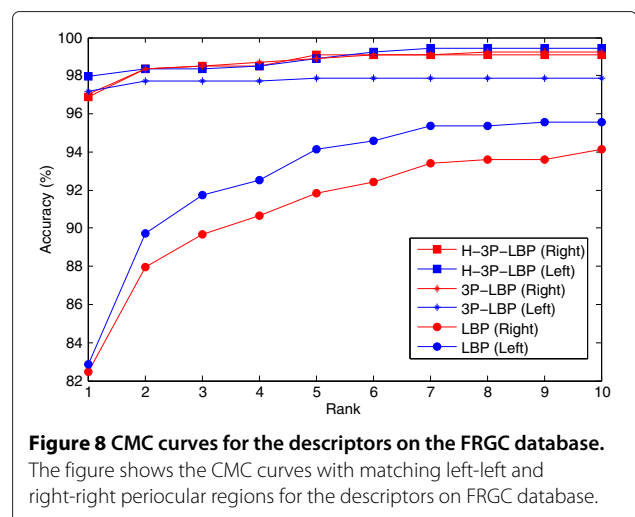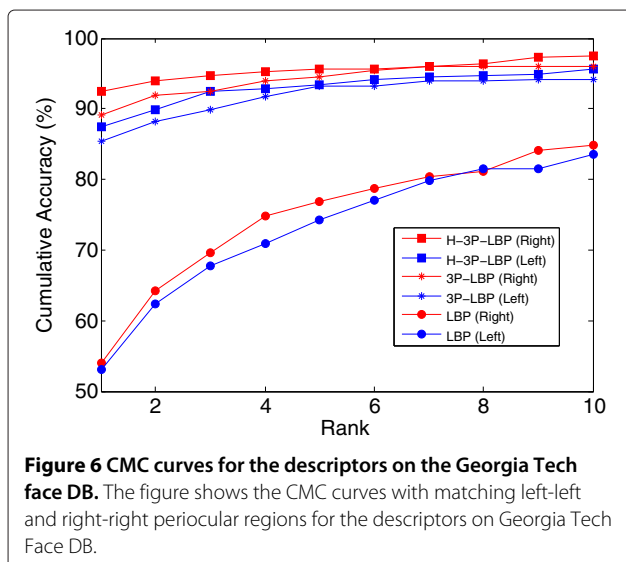
Figure 6 CMC curves for the descriptors on the Georgia Tech face DB. The figure shows the CMC curves with matching left-left and right-right periocular regions for the descriptors on Georgia Tech Face DB.

Figure 8 CMC curves for the descriptors on the FRGC database. The figure shows the CMC curves with matching left-left and right-right periocular regions for the descriptors on FRGC database.

Figure 9 Sample periocular images showing various challenges. The figure shows sample images from the Georgia Tech face database and the MORPH Album 1. Each row shows examples of occlusions, pose variations, eyeglasses, and closed eyelids, respectively.

FaceVACS V8.5 [26]. The face region in the image is geometrically normalized by aligning the images based on the eye center coordinates. We follow a procedure similar to that of [27] to align the images. The alignment process involves scaling, rotation, and cropping the face region to a specified size, such that the eye centers are horizontally aligned and placed on standard pixel locations. Figure 4 shows the entire image alignment process for a sample image.

3.3 Periocular region segmentation

The periocular region is extracted from the aligned face image prior to the feature descriptor computation. There are no standard guidelines in existing literature that clearly define the periocular region. Often, the periocular region is defined as the skin region around the eyes, the eyes, and the eyebrows. The eyebrows are generally included in the periocular region since it helps in discrimination between subjects. The region as defined above is more accurately

Table 5 Rank-1 accuracies for neutral-neutral (gallery-probe) matching for Georgia Tech face database

	Left (%)	Right (%)	Fusion (%)
LBP	49.27	51.38	*62.64*
3P-LBP	80.92	85.32	*86.30*
H-3P-LBP	82.06	87.60	*88.26*

Values in italics are the best achieved results.

Table 6 Rank-1 accuracies for neutral-smile (gallery-probe) matching for Georgia Tech database

	Left (%)	Right (%)	Fusion (%)
LBP	46.72	54.09	*64.75*
3P-LBP	74.59	80.33	*81.97*
H-3P-LBP	77.05	82.79	*82.79*

Values in italics are the best achieved results.

Table 7 Rank-1 accuracies for smile-smile (gallery-probe) matching for Georgia Tech database

	Left (%)	Right (%)	Fusion (%)
LBP	45.83	37.96	44.44
3P-LBP	60.65	57.87	*61.57*
H-3P-LBP	60.65	57.41	*64.35*

Values in italics are the best achieved results.

Table 9 Rank-1 accuracies for non-frontal - frontal (gallery-probe) matching for Georgia Tech database

	Left (%)	Right (%)	Fusion (%)
LBP	44.92	51.15	*56.72*
3P-LBP	82.95	82.29	*85.57*
H-3P-LBP	83.93	87.21	*89.51*

Values in italics are the best achieved results.

known as the periorbital region, which relates to the bony structure of the eye orbit and the soft tissue around this structure. Periocular correctly refers to the soft tissue of the region internal to the eye orbit. However, for this work, we adopt the periocular term currently used in the literature [1]. In this work, the segmentation of the periocular region includes the eyes, eyebrows, and the skin region around the eyes. We perform an automatic segmentation using the coordinates of the eye centers in the aligned image. The automatic segmentation is feasible due to the placement of the eye centers on standard pixel locations during the alignment process. A region of size 128×128 pixels centered by the eye center is extracted for both the left and right eye region from the aligned image. It is to be noted that the iris is not masked from the extracted periocular images, which can have some effect on the recognition performance. Some researchers have chosen to mask the eyeball area and utilized information from the shape of the eye and the eyebrow region. However, the surface level texture of the iris can provide additional cues and, hence, can help improve the recognition accuracy. Hence, in this work, we match against both open and closed eyes.

3.4 Effect of periocular image size vs. the number of image scale

This experiment was designed to analyze the effect of the extracted periocular image size and the number of image levels in the H-3P-LBP computation on the recognition performance. The frontal and neutral expression images from the ND Twins database were utilized for this experiment. Choi et al. [28] have shown that an inter-pupillary distance (IPD) of at least 60 pixels is required for successful recognition. The IPD varies with the image size, and hence, variations in the image size can significantly affect the recognition performance. It is to be noted that the IPD is varied in our experiments by varying the size of the extracted periocular region individually rather than resizing the aligned full-face image. In addition to the periocular image size, the number of scales in the image pyramid computed for the H-3P-LBP can have an impact on the recognition performance as images of different sizes are considered at each level of the pyramid.

The periocular images of a subject from the ND Twins database are equally split into gallery and probe by random bootstrap sampling and are used in the recognition process. This process is repeated three times for each combination of image size and number of image scales. Table 2 lists the rank-1 recognition accuracies for various image scales and image sizes. First, it is to be noted that there is an insignificant effect of the number of levels in the hierarchy when the image is either in its original dimensions or enlarged. However, the performance significantly reduces with the reduction of the image size to a lower dimension than the original size. This indicates that texture information is lost during the reduction. This in turn suggests that the recognition rate improves for images with an IPD of at least 60 pixels when compared with images with an IPD of less than 60 pixels.

3.5 Recognition accuracy

The periocular recognition performance was studied using the datasets described in Section 3.1. A closed set identification was performed for all the experiments; hence, no subject was considered an impostor during recognition. Each dataset was divided equally into gallery and probe, where the images for the gallery and probe for a subject were randomly selected. Every probe image was compared against all the gallery images using uniform LBP, 3P-LBP, and H-3P-LBP matching techniques. The

Table 8 Rank-1 accuracies for frontal - non-frontal (gallery-probe) matching for Georgia Tech database

	Left (%)	Right (%)	Fusion (%)
LBP	36.65	44.34	*51.36*
3P-LBP	74.43	80.77	*84.16*
H-3P-LBP	76.47	83.94	*87.78*

Values in italics are the best achieved results.

Table 10 Rank-1 accuracies for non-frontal - non-frontal (gallery-probe) matching for Georgia Tech database

	Left (%)	Right (%)	Fusion (%)
LBP	47.48	39.54	*50.22*
3P-LBP	70.56	69.41	*77.06*
H-3P-LBP	74.03	72.29	*77.49*

Values in italics are the best achieved results.

Table 11 Rank-1 accuracies obtained with age-varying data from MORPH Album 1

	Left (%)	Right (%)	Fusion (%)
LBP	2.32	2.89	*3.96*
3P-LBP	19.21	24.04	*26.93*
H-3P-LBP	21.04	27.51	*28.96*

The table indicates the rank-1 accuracies obtained with age gap (up to 20 years) between the gallery and probe (from MORPH Album 1) for both the left and right periocular regions for LBP, 3P-LBP, and H-3P-LBP descriptors. Values in italics are the best achieved results.

Table 12 Rank-1 accuracies obtained using eyelid closed images from Georgia Tech database

	Left (%)	Right (%)	Fusion (%)
LBP	62.16	65.76	*71.17*
3P-LBP	86.49	85.59	*87.39*
H-3P-LBP	*87.39*	84.69	86.49

Values in italics are the best achieved results.

results of the experiments are provided in terms of cumulative match characteristic (CMC) curves and in terms of the rank-1 recognition accuracy. Matching was performed for the left-left and right-right gallery-probe periocular image pair as previous research has indicated that the left and right regions are sufficiently different. The left and right periocular regions were determined based on the location of the nose with respect to the inner corner of the eye. In other words, the left and right periocular regions were defined from the subject's perspective.

It is to be noted that the gallery and the probe include images containing variations in pose, expressions, and illumination, besides neutral expression and frontal pose. All the datasets include images with the presence of facial hair and glasses, with the exception of the Twins database. The aforementioned experiment can be understood as a baseline for matching under non-ideal conditions. Table 3 indicates the rank-1 accuracies obtained for various descriptors on all the datasets. Table 4 lists the rank-1 accuracies from the score-level fusion approach. Figures 5, 6, 7, and 8 show the CMC curves for all the descriptors on MORPH, Georgia Tech, the Twins database, and FRGC, respectively.

From these results, it can be noticed that patch-based approaches perform better when compared with the pixel-based computation of LBP. The performance of all the

descriptors on MORPH Album 1 indicates the significance of effects such as template aging, pose variations, expression changes, etc. in periocular recognition. Also, the images of MORPH are scanned photographs in contrast with the other two datasets. This indicates the need for better-matching algorithms in case of recognizing subjects using scanned low-resolution images. It is also to be noted that there is a significant difference with the recognition accuracies for the left and right periocular regions, which indicates the profile-specific features that are extracted by the descriptors. The matching accuracies indicate that the performance was improved by computing the 3P-LBP in a hierarchical fashion. This is due to the extraction of micro- and macro-patterns, both of which are required for better texture discrimination. The best recognition performance was achieved when the left and right periocular scores are fused together, which indicates the use of side information and side-specific features for better recognition.

3.6 Recognition under non-ideal conditions

Generally, variations in pose, illumination, and expressions are considered as challenges by the face recognition community. Figure 9 contains sample periocular images under various challenges. In this section, we present a discussion on the effect of these challenges in periocular-based recognition.
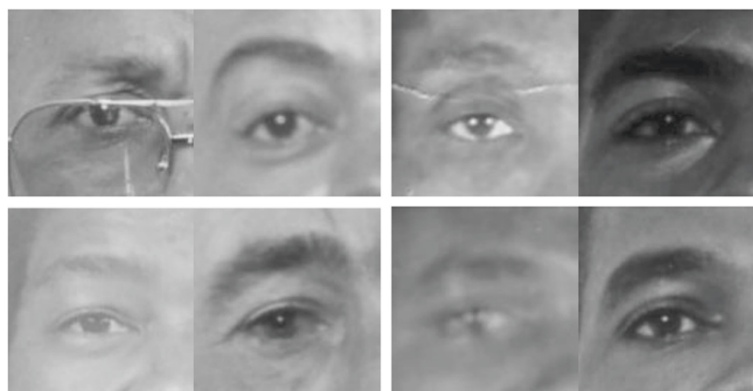

Figure 10 Examples of failure cases from MORPH Album 1. The figure shows sample gallery-probe image pairs from the MORPH Album 1 that were incorrectly recognized by the recognition framework.

Table 13 Rank-1 accuracies obtained with images from Georgia Tech database and ND Twins database with eyeglasses

Approaches	Georgia Tech			ND Twins		
	Left (%)	Right (%)	Fusion (%)	Left (%)	Right (%)	Fusion (%)
LBP	26.58	24.68	*31.65*	13.33	10.00	*16.67*
3P-LBP	72.15	63.92	*79.75*	46.67	20.00	*40.00*
H-3P-LBP	77.22	70.25	*81.01*	53.33	30.00	*56.67*

Values in italics are the best achieved results.

The performance of all the descriptors was analyzed from the perspective of matching gallery and probe images with the following scenarios: (1) neutral-neutral (expressions), (2) neutral-smile, (3) smile-smile, (4) frontal pose - non-frontal pose, (5) non-frontal pose - frontal pose, (6) non-frontal pose - non-frontal pose, (7) no glasses-with glasses, and (8) eyes open-eyes closed. In addition, the effect of template aging on periocular recognition was also studied. The Georgia Tech face database and the MORPH Album 1 were utilized for this study. Recognition experiments were conducted using the images that were categorized based on the presence of the above mentioned factors.

For the neutral-neutral gallery-probe scenario, one image from each subject was used in the gallery, and the remaining images were used as probe. The experiment studying the effect of template aging utilized the youngest images as gallery and the older images as probe. This can be correlated with the real-world scenario of verifying passports or in security, where an already-enrolled young-aged image is compared against the later aged one. For the remaining experiments, all the images from the respective subsets were used as gallery and probe for each scenario.

3.6.1 Expression variations

Images from the Georgia Tech face database were used for these experiments. The effect of change in expressions in the periocular region was analyzed by comparing the neutral expression image with those having a smiling expression. The database included 613 images from 50 subjects with neutral expression and 122 images from 39 subjects with a smiling expression.

Tables 5, 6, and 7 show the rank-1 accuracies for the scenarios neutral-neutral, neutral-smile, and smile-smile. It can be noticed that change in facial expression results in degradation of the recognition performance for patch-based LBP approaches when compared with matching against images with neutral expression. The generation of wrinkles near the eye and the rising or lowering of the eyebrows while exhibiting the expression causes changes to the periocular region. Also, it is to be noted that there is a performance degradation for the smile-smile scenario when compared with the neutral-neutral and neutral-smile scenarios. This is due to the insufficient set of

gallery images that could span the entire set of expression changes. This suggests the use of images with neutral expressions for recognition task. Also, it is evident that the eyebrows act as a discriminative cue between individuals. In contrast, LBP shows an improved performance for the right periocular region and the score-level fusion approach. One reasoning could be that the pixel-based computation of LBP captures these variations minimally when compared with the patch-based computation of LBP.

3.6.2 Pose variations

In real-world scenarios such as video surveillance, the pose of the facial images is not always frontal. This also introduces a pose variation with the periocular region of the face and occlusions depending on the pose of the face. We collected 305 frontal pose images and 442 images with non-frontal pose from the Georgia Tech database. Both subsets include images from each subject in the database, and each subset plays the role of gallery as well as probe. Tables 8, 9, and 10 show the rank-1 recognition accuracy for (1) frontal to non-frontal, (2) non-frontal to frontal, and (3) non-frontal to non-frontal scenarios. It can be seen that large variations in pose of the face significantly affects performance. This is particularly evident from the performance on the non-frontal to non-frontal scenario, where the left and right periocular regions achieve different recognition rates due to variation in the pose between them. This performance degradation is similar to the performance decline that is seen with traditional face recognition and can be viewed as occlusion and skew. As the face is turned, out-of-plane rotation with the image sensor, parts of the periocular region towards the direction of rotation becomes occluded while the other side of the face undergoes skew (elongation and perspective change).

3.6.3 Template aging

To study the effect of time lapse between the gallery and the probe in the recognition performance, an experiment utilizing the images from MORPH Album 1 was performed. The images of a subject was near equally divided into gallery and probe, where the gallery included relatively younger images of the subject and the probe included the more recent images of the subject.

Figure 11 Examples of failure cases from Georgia Tech database. The figure shows sample gallery-probe image pairs from Georgia Tech DB that were incorrectly recognized by the recognition framework.

Table 11 shows the rank-1 recognition accuracies obtained from this experiment. As expected, the results show a significant effect of template aging in recognition performance. Comparing the results with the baseline performance in Table 3, it can be seen that larger age gaps between the probe and the gallery result in larger dissimilarities in facial features and hence the failure of the LBP and its variants to effectively capture the intra-class similarities and the inter-class dissimilarities. Figure 10 shows some examples of mismatches from the experiment. It is to be noted that the difficulty arises with the presence of image artifacts, presence of glasses, pose variations, and nonetheless, age gap between probe and gallery.

3.6.4 Effect of closed eyelids
The iris and the eye region are a soft texture that can provide additional information on the periocular region and thus can improve the recognition performance. However, real-world images are captured under non-ideal conditions where motion of the iris and the eyelid is possible, which causes masking of the eye region. Although, closing of the eyelid can be considered as masking of the eye region, it also provides some additional texture information, which can aid the recognition process. To study the effect of the presence of this additional texture and the masking of the eye region, we conducted experiments where the gallery involved images with the eyes open, and the probe involved images with the eyelid closed. We collected 111 images with the eyelid closed from the Georgia Tech database as the probe. The remaining images from the database were included in the gallery. Table 12 shows the rank-1 recognition accuracies for the left, right, and the score-level fusion. Although, we observed degradation in performance for the patch-based approaches, there is an improvement in the performance for the pixel-based computation of

LBP. This possibly indicates the advantage of using pixel-based computation of local features under such non-ideal conditions.

3.6.5 Effect of eyeglasses
The presence of eyeglasses on the face of a subject could hide a significant portion of the periocular region. Eyeglasses have been shown as discriminative cues in the face verification task [29,30]. To validate prior work for periocular recognition, recognition was performed using images with eyeglasses as probe and the images without eyeglasses as the gallery. Table 13 shows the rank-1 recognition accuracies obtained from uniform LBP and its variants on the Georgia Tech and the ND Twins datasets. It is to be noted that the recognition performance for all the approaches degrades when compared with the baseline. Since the eyeglasses can be treated as occlusions/disguise of the periocular region, it can be deduced that eyeglasses may not be a suitable feature for effective discrimination between subjects. Figure 11 shows examples of incorrectly recognized gallery-probe image pairs by the proposed framework. It can be seen that the presence of glasses and large pose variations affect the performance of the system.

4 Conclusions and future work
In this paper, we investigated the performance of LBP and its variants in periocular-based recognition using unconstrained face images. We proposed the multi-scale, hierarchical three-patch LBP framework, which is a variant of the three-patch LBP. The matching performance was evaluated using the uniform LBP, three-patch LBP, and the hierarchical three-patch LBP. The effects of covariates such as pose variations, facial expression, template aging, and occlusions on periocular recognition performance were discussed. Experiments on four challenging datasets yield the best recognition results for the proposed

method when compared with LBP and its variants. Experiments indicate that the best results were achieved when matching was performed for both the left and right periocular regions individually and then fusing their scores. The results also indicate that there is significant discrimination between the left and right periocular region of the same subject. The performance of the patch-based LBPs can be improved when images with neutral expressions are used. However, the uniform LBP is significantly robust for both neutral and varied expressions on the face.

There is a significant effect on the recognition performance due to large pose variations, while the effects of minimal pose variations are insignificant due to the pose-invariant nature of the LBP operator. Aging effects are prominent in the periocular region of a face, which increases the intra-class dissimilarities as the age gap between the gallery and probe increases. Our experiments also indicate that conventional LBP and its variants fail to capture these age-based differences.

Masking of the iris and the eye region has an impact on the performance of patch-based descriptors, while it improves the performance of pixel-based LBP. While the presence of eyeglasses help face recognition systems, they degrade the performance of a periocular recognition system. The performance of periocular recognition could be further improved with the consideration of cues such as eyelashes, eye shape, and size.

In future work, we will explore the use of different distance measures for matching as the Euclidean distance has been proven to not be the most robust in the face domain. Furthermore, we will explore developing texture-based features that are resilient to aging changes. Developing a texture-based age-invariant texture technique would have far-reaching and sweeping impacts on face-based biometric techniques.

Competing interests

Acknowledgements

References

1. U Park, A Ross, AK Jain, in *Proceedings of the 3rd IEEE International Conference on Biometrics: Theory, Applications, and Systems*. Periocular biometrics in the visible spectrum: a feasibility study (IEEE Piscataway, 2009), pp. 153–158
2. S Bharadwaj, HS Bhatt, M Vatsa, R Singh, in *Proceedings of the 4th IEEE International Conference on Biometrics: Theory, Applications, and Systems*. Periocular biometrics: when iris recognition fails (IEEE Piscataway, 2010), pp. 1–6
3. D Woodard, S Pundlik, P Miller, R Jillela, A Ross, in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*. On the fusion of periocular and iris biometrics in non-ideal imagery (IEEE New York, 2010)
4. P Miller, A Rawls, S Pundlik, D Woodard, in *Proceedings of the 2010 ACM Symposium on Applied Computing*. Personal identification using periocular skin texture (ACM New York, 2010)
5. D Woodard, S Pundlik, J Lyle, P Miller, in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Periocular region appearance cues for biometric identification (IEEE New York, 2010)
6. JR Lyle, PE Miller, SJ Pundlik, DL Woodard, Soft biometric classification using local appearance periocular region features. Pattern Recognit. **45**(11), 3877–3885 (2012)
7. CN Padole, H Proenca, in *The 5th IAPR International Conference on Biometrics (ICB)*. Periocular recognition: analysis of performance degradation factors (IEEE New York, 2012), pp. 439–445
8. FJ Xu, K Luu, M Savvides, TD Bui, CY Suen, in *International Joint Conference on Biometrics (IJCB)*. Investigating age invariant face recognition based on periocular biometrics (IEEE New York, 2011)
9. Y Dong, DL Woodard, in *International Joint Conference on Biometrics (IJCB)*. Eyebrow shape-based features for biometric recognition and gender classification: a feasibility study (IEEE New York, 2011), pp. 1–8
10. KP Hollingsworth, KW Bowyer, PJ Flynn, Useful features for human verification in near-infrared periocular images. Image Vis. Comput. **29**(11), 707–715 (2011)
11. KP Hollingsworth, SS Darnell, PE Miller, DL Woodard, KW Bowyer, PJ Flynn, Human and machine performance on periocular biometrics under near-infrared light and visible light. IEEE Trans. Inf. Forensics Secur. **7**(2), 588–601 (2012)
12. AV Nefian, MH Hayes, in *International Conference on Image Processing*. Maximum likelihood training of the embedded HMM for face detection and recognition (IEEE New York, 2000)
13. PJ Philips, PJ Flynn, KW Bowyer, RWV Bruegge, PJ Grother, GW Quinn, M Pruitt, in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. Distinguishing identical twins by face recognition (IEEE New York, 2011), pp. 15–192
14. PJ Phillips, PJ Flynn, T Scruggs, KW Bowyer, W Worek, in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*. Preliminary face recognition grand challenge results (IEEE New York, 2006), pp. 15–24
15. K Ricanek Jr, T Tesafaye, in *Proceedings of the 7th International Conference on Automatic Face Gesture Recognition*. MORPH: A longitudinal image database of normal adult age-progression (IEEE New York, 2006), pp. 341–345
16. L Chen, H Man, AV Nefian, Face recognition based multi-class mapping of Fisher scores. Pattern Recognit., Spec. Issue Image Underst. Digital Photographs. **38**(6), 799–811 (2005)
17. THN Le, K Luu, K Seshadri, M Savvides, in *IEEE 5th International Conference on Biometrics: Theory, Applications, and Systems*. A facial aging approach to identification of identical twins (IEEE Piscataway, 2012)
18. AM Albert, K Ricanek Jr, The MORPH database: investigating the effects of adult craniofacial aging on automated face-recognition technology. Forensic Sci. Commun. **10**(2) (2008)
19. T Maenpaa, T Ojala, M Pietikainen, M Sori-ano, in *Proceedings 15th International Conference on Pattern Recognition*. Robust texture classification by subsets of local binary patterns (IEEE New York, 2000)
20. L Wolf, T Hassner, Y Taigman, in *Faces in Real-Life Images Workshop in ECCV*. Descriptor based methods in the wild (2008)
21. E Shechtman, M Irani, in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. Matching local self-similarities across images and videos (IEEE New York, 2007), pp. 1–8
22. SW Lee, SZ Li, in *LNCS 4642*, ICB'07. Learning multi-scale block local binary patterns for face recognition (Springer Berlin, 2007), pp. 828–837
23. D Woodard, S Pundlik, J Lyle, P Miller, in *Computer Vision and Pattern Recognition Workshops*. Periocular region appearance cues for biometric identification (IEEE New York, 2010), pp. 162–169
24. K Ricanek, E Boon, in *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks (IJCNN)*. The effect of normal adult aging on standard PCA face recognition accuracy rates (IEEE New York, 2005), pp. 2018–2023
25. K Ricanek, E Boone, E Patterson, in *Proceedings of the 6th IASTED Visualization, Imaging, and Image Processing (VIIP)*. Craniofacial aging on the eigenface biometric (Palma de Mallorca, 20–30 August 2006), pp. 249–253

26. KitCognitecSystemsGmbH FaceVACS Software Developer (2012). http://www.cognitec-systems.de
27. PJ Philips, JR Beveridge, B Draper, G Givens, A O'Toole, D Bolme, J Dunlop, YM Lui, H Sahibzada, S Weimer, An introduction to the good, the bad, and the ugly face recognition challenge problem. Image Vis. Comput. **30**(3), 206–216 (2012)
28. HC Choi, U Park, AK Jain, PTZ camera assisted face acquisition, tracking and recognition. Biometrics: Theory, Appl., Syst., 1–6 (2010)
29. H Ling, S Soatto, N Ramanathan, D Jacobs, Face verification across age progression using discriminative methods. IEEE Trans. Inf. Forensics Secur. **5**, 82–91 (2010)
30. G Mahalingam, C Kambhamettu, in *Proceedings of the 7th Indian Conference on Computer Vision, Graphics and Image Processing*. Face verification across age progression using AdaBoost and local binary patterns (ACM New York, 2010), pp. 101–108