UWA Research Publication

Periocular Biometric Recognition using Image Sets

Muhammad Uzair, Arif Mahmood, Ajmal Mian, Chris McDonald
Computer Science & Software Engineering
The University of Western Australia
35 Stirling Highway, Crawley, WA, Australia
{uzair, arifm, ajmal, chris}@csse.uwa.edu.au

Abstract

Human identification based on iris biometrics requires high resolution iris images of a cooperative subject. Such images cannot be obtained in non-intrusive applications such as surveillance. However, the full region around the eye, known as the periocular region, can be acquired non-intrusively and used as a biometric. In this paper we investigate the use of periocular region for person identification. Current techniques have focused on choosing a single best frame, mostly manually, for matching. In contrast, we formulate, for the first time, person identification based on periocular regions as an image set classification problem. We generate periocular region image sets from the Multi Biometric Grand Challenge (MBGC) NIR videos. Periocular regions of the right eyes are mirrored and combined with those of the left eyes to form an image set. Each image set contains periocular regions of a single subject. For image-set classification, we use six state-of-the-art techniques and report their comparative recognition and verification performances. Our results show that image sets of periocular regions achieve significantly higher recognition rates than currently reported in the literature for the same database.

1. Introduction

Human identification based on iris biometrics has been well investigated [3, 11, 10]. Such an identification is based on high resolution iris images often acquired by iris scanners [14, 15]. Iris scanning requires the subjects to open their eyes fully and look into the scanner which may not be socially acceptable to users. In many practical scenarios, the subject may not be cooperative, willing, or even aware that he/she is being imaged. For example, in banks and shopping malls, customers may need to be identified without causing inconvenience to them. Other applications include security and surveillance where cameras are often located at large distances, and hence, the resolution of the iris is too low to be used as a stand-alone biometric.

Figure 1. Some example applications in which periocular biometrics may be more effective than the full face biometrics.

Recently, researchers have used the area around the eye, the periocular region, as a stand-alone biometric with promising results [14, 15]. Periocular region based human identification offers advantages over full face biometrics as it is least affected by expression variations, aging effects [7] and the changes due to growth of male facial hair. Moreover, full face recognition performance degrades in the presence of pose variations whereas the periocular region based identification may perform better in the case of extreme pose changes when only one eye is completely visible. In these cases, the visible periocular region may be mirrored and matched to the opposite side periocular region in the database. Finally, periocular region based recognition will remain effective even if most of the lower face region is occluded and as long as only one eye remains visible. Fig. 1 shows some examples.

Most of the previous work on periocular biometric recognition is based on single image matching [5, 20, 21, 2, 22]. In many cases, a single best frame per subject is manually selected and placed in the gallery. Each test image in the query sequence is matched to the gallery images to find the best match. Such techniques inherently suffer from the lack of information since only one image cannot contain all possible variations in the periocular region of an individual

(see Fig. 2). As an example, if the best image is taken from a frontal view with centered eyeballs and the query images have some pose variations with eyeballs at the extreme ends, recognition performance will significantly degrade.

In contrast to the existing single image based approaches, we formulate periocular region based person identification as an image set classification problem. Each set contains multiple periocular region images of the same person and represent a wide range of variations in the periocular region. These variations include different eyelid positions, different illuminations, eyeball movements and pose variations. Multiple images in the same set may compliment such appearance variations of the periocular region. We construct a gallery from image sets of many identities. The query set also contains multiple periocular region images of the same person and is assigned the label of the nearest gallery set. Compared to single image matching, set-to-set matching offers significantly more information. Although, the left and right periocular regions are different, we mirror the right periocular regions and combine them with the left periocular regions to form a single set per identity. We believe that this strategy better suits linear modeling techniques.

For classification, we use six existing state-of-the-art image set classification techniques and compare their performances with the exhaustive nearest neighbour approach. These techniques include the Affine Hull based Image Set Distance (AHISD) [4], Convex Hull based Image Set Distance (CHISD) [4], Discriminative Canonical Correlation (DCC) [9], Manifold-Manifold Distance (MMD) [19], Manifold Discriminant Analysis (MDA) [18], and Sparse Approximated Nearest Point (SANP) distance [6]. A briefly overview of these techniques is given in Section 2.

We performed extensive experiments on the Multi Biometric Grand Challenge (MBGC) NIR video dataset [16, 1]. The best image set based classification is observed for the SANP algorithm [6] which has a 97.70% recognition rate. A rank four recognition rate $\geq$ 99.00% is observed for AHISD, CHISD, MDA, and SANP. This is significantly higher than the single image based recognition rates already reported in the literature for the same dataset.

2. Related Work on Periocular Biometrics

Periocular biometric refers to the face region surrounding the eye. It contains iris [15], eyes, eyelids, eye lashes, and part of the eyebrows [20]. Compared to face and iris biometrics, human recognition using periocular biometrics has not been well investigated. Initial feasibility studies of periocular biometrics have been done by Park et. al [15] and extended in [14]. They detected iris in the visible spectrum images and extracted features at fixed grid positions around the center of the limbus circle. Their method relies on accurate iris detection and is vulnerable to eyeball movements.

Figure 2. Variations in the appearance of the periocular region of the same subject.

Some researchers have investigated the use of Local Binary Pattern (LBP) features [17] to represent the texture of the periocular region for recognition purpose. Miller et al. [13] used city block distance to classify LBP features of the periocular region. Their work was extended by Adams et al. [2] using genetic algorithms to select the optimal subset of LBP features for the matching of periocular regions. Woodard et al. [20] used the LBP features and color histograms to represent the local appearance of the periocular region. They observed better recognition performance of city block distance for matching LBP features and Bhattacharya coefficient for matching color histograms.

Xu et al. [22] observed improvements in recognition rates by combining LBP with other features including DCT, DWT, Gabor filters, LoG filters, Walsh transform [8], SIFT, and SURF. Woodard et al. [21] simultaneously used the iris and periocular biometrics by performing score-level fusion. LBP features were extracted from the periocular region while the iris texture was encoded using the Gabor filters. Their experiments showed that the periocular recognition performed better than iris recognition for the MBGC NIR portal videos.

Most of the existing research on the periocular biometrics has mostly investigated the use of texture features calculated from a single best periocular image. Therefore, these methods cannot efficiently handle the periocular region variations including the eyeball and eyelid movements in frontal poses under varying illumination conditions. If pose variations are also introduced, these techniques may observe further performance degradation.

In order to mitigate some of these challenges, we propose image set based approach for periocular biometric recognition. For each subject, the gallery may contain one or more image sets and each image set may contain multiple images of the same subject complementing a wide range of periocular region variations. Due to the availability of significantly more information, in our experiments, we observe much better recognition results compared to the single image based results already reported in the literature.

3. Image Set Classification

Image set classification algorithms have been well investigated for the face biometrics. However, to the best of

our knowledge, periocular region based person identification has not been formulated as an image set classification problem. We believe that set based classification is more suitable for periocular region biometrics. In this paper, we use six state-of-the-art image set classification algorithms for periocular region based human recognition. These techniques may be broadly divided into two categories namely, sample based and the structure based techniques. In the following sections, both categories are explained.

3.1. Sample Based Image Set Classification

Sample based techniques measure the distance between nearest neighbour samples of two image sets. Let $X = \{x_i\}_{i=1}^n \in \mathcal{R}^{m \times n}$ be an image set, where $x_i \in \mathcal{R}^m$ is a feature vector and n be the number of feature vectors in a set which may vary across the image sets. The feature vectors can simply be the image pixel values or some features calculated from the pixels such as the PCA (Principal Component Analysis) coefficients or LBP features [17]. Each image set may be considered as a point cloud in $\mathcal{R}^m$. All points in a probe image set X_p are compared with all points in each gallery set X_g to find the nearest pair of points (x_i, x_j) such that $x_i \in X_p$ and $x_j \in X_g$. If x_i and x_j have zero mean and unit magnitude, the nearest neighbour pair (x_i, x_j) is the one that maximizes the cosine of the angular distance:

$$\max_g \big(\max_{x_i, x_j} X_g^t X_p\big). \qquad (1)$$

The probe image set label is predicted as the label of x_j.

Cevikalp and Triggs [4] considered each image set as a convex geometric region in $\mathcal{R}^m$. Set dissimilarity was measured by the distance of closest approach between the regions represented by the affine (AHISD) or convex hulls (CHISD). The region contains all the affine combinations $x_p = \sum_{i=1}^n X_{pi}\alpha_{pi}$ and $x_g = \sum_{j=1}^n X_{gj}\alpha_{gj}$, where $\alpha_p, \alpha_g \in \mathcal{R}^m$. For the case of affine hull $\sum_{i=1}^n \alpha_{pi} = \sum_{j=1}^n \alpha_{gj} = 1$ and for the convex hull $0 \leq (\alpha_{pi}, \alpha_{gj}) \leq 1$. The minimum distance was computed as

$$\min_g \big(\min_{\alpha_p, \alpha_g} ||X_g \alpha_g - X_p \alpha_p||^2\big). \qquad (2)$$

For the case of affine hull, the minimum distance was computed using least squares while for the case of convex hull, an SVM was trained to separate probe and gallery sets belonging to different classes.

Instead of searching the nearest points with dense combinations of samples in the corresponding image set, Hu et al. [6] proposed that each of the two points should be able to be approximated by a sparse combination from the samples of the respective set. They argued that the sparse approximated nearest points (SANP) will lie close to some facet of the affine hull and hence, implicitly incorporate structural information of the sets as well. By restricting the SANPs to be close the facets, this approach can reject outliers resulting in more accurate classification.

Figure 3. Sample images from one image set.

3.2. Structure Based Image Set Classification

Structure based techniques represent the underlying structure of a image set with one or more linear subspaces. Structural similarity of the sets is usually measured using subspace to subspace distance. Kim et al. [9] proposed Discriminative Canonical Correlation (DCC) which performs discriminative learning using canonical correlations. More specifically, a discriminant function was learned that maximizes the within-class and minimizes the between-class canonical correlations. Image sets were transformed optimally by this discriminant function and compared using their canonical correlations.

Wang et al. [19] proposed Manifold-Manifold Distance (MMD) which clusters each image set into multiple linear subspaces called local models. The similarity between two sets is calculated by computing the subspace to subspace distance between the nearest local models. Subspace to subspace distance is based on canonical correlations. However, the nearest point distance is also combined with the structural similarity to calculate the final similarity between two sets. Wang and Chen [18] proposed Manifold Discriminant Analysis (MDA) that uses a hierarchical divisive clustering approach to construct local linear models of each set. The local models are transformed by a linear discriminant function where different classes are better separable. The similarity between two sets is calculated as the pair-wise local model distances in the learned embedding space.

4. Image Sets and Feature Extraction

We have performed extensive experimentation on periocular images extracted from the NIR face videos of the MBGC portal challenge dataset version 2 [1]. MBGC is a challenging dataset with significant illumination variations, motion blur, and specular reflections. Our experiment consists of 85 subjects from the MBGC dataset. Fig. 3 shows sample periocular images from a set. Each subject has 2 to 7 image sets and each image set contains 9 to 27 images. The gallery is constructed by randomly selecting one image-set for each of the 85 subjects. The remaining 132 image-sets are used as probes. The experiments are repeated three folds
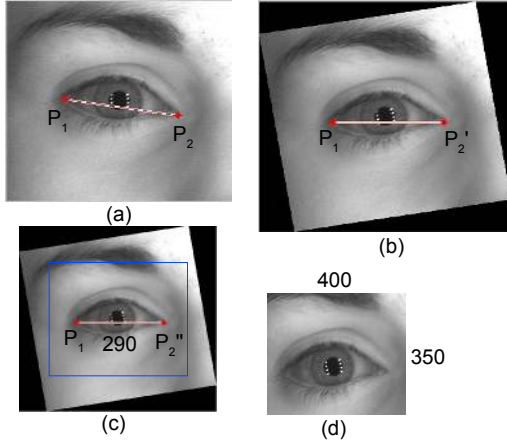
(a)

(b)

400

350

P_1 290 P_2''

(c)

(d)

Figure 4. Periocular region normalization

Right Left Right flipped Left

Figure 5. Mirroring the periocular images to normalized form.

by randomly selecting different gallery sets each time.

4.1. Preprocessing

The MBGC portal challenge consists of NIR videos of the faces captured while people walk through a portal towards the camera. Due to varying distance from the camera, significant scale changes occur across consecutive frames in the same video. Moreover, due to head pose variations, some rotational effects are also observed across different videos. These scale and rotation variations do not provide any discriminative information. To avoid the image set classification algorithms to model these undesirable variations, we have performed rotation and scale normalization.

In each video frame, we manually marked two eye corners P_1 and P_2 on each eye and cropped a periocular region of size 600×600 pixels. The cropped image is rotated such that the line through P_1 and P_2 becomes horizontal. Then the image is scaled keeping a fixed distance of 290 pixels between P_1 and P_2. An image of size 400×350 pixels is then cropped. Fig. 4 illustrates the normalization process. The right side periocular regions are mirrored to obtain images similar to the left side region (Fig. 5). This mirroring process may also be considered as normalization because the resulting images contain only the appearance variations of the left periocular region and allows for better modeling. We consider the resulting images in canonical form, only containing the required discriminative information. The gallery and probe image sets used in our experiments consist of these normalized periocular images.

Frames containing very low information (e.g. dark and blurred) or very few frames are discarded. Performance of periocular biometric degrades in the presence of strong mascara eyelash extension. To keep our experiments simple, we do not use the videos where a strong mascara eyelash extension is used by the subject. Our final experimental data consists of 217 image sets and 3163 periocular images of 85 subjects. To assist other researchers, the normaliza-
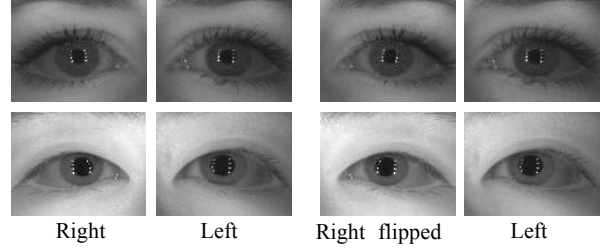
tion code along with eye corner points and the exact video sequences used in our experiments will be made publicly available once the paper is published.

4.2. Feature Extraction

In our experiments, we have used four different types of feature vectors with varying amount of information. Note that for PCA, only the gallery sets are used as training data.

1. **Raw Pixel Values:** The normalized periocular images are scaled down to 20×18 pixels and the pixel values are used as features.

2. **LBP Features:** For the 20×18 pixels images, LBP features are extracted using circular (8,1) neighborhoods [17]. The resulting LBP coded images are used as feature vectors.

3. **PCA Dimensionality Reduction:** The dimensionality of the normalized images is reduced by projecting them on 400 most significant PCA basis. The resulting 400 coefficients are used as feature vectors.

4. **LBP Features Followed by PCA:** The LBP features are computed over the normalized images and the dimensionality of the LBP codes is reduced to 400 dimensions by applying PCA.

5. Experimental Setup

The performance of the periocular biometrics is tested using six state-of-the-art image-set classification algorithms namely, Affine and Convex Hull based Image Set Distance (AHISD and CHISD) [4], Sparse Approximated Nearest Points (SANP) distance [6], Discriminative Canonical Correlation [9], Manifold-Manifold Distance [19] and Manifold Discriminant Analysis [18]. Additionally, results are also reported for the simple nearest neighbour (NN) technique based on Eq. (1). For every algorithm, we performed three fold experiments for each of the four feature vector types.

For the sample based algorithms, the default input parameter values are used. For the structure based algorithms, the required parameters are carefully tuned so that the best performance can be obtained. For DCC, the subspace dimensions are set to 10 which preserves 90% energy and the corresponding 10 maximum canonical correlations are used
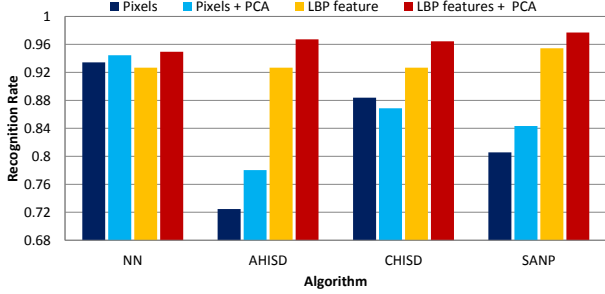
Figure 6. Recognition Rates averaged over three folds for the sample based image set classification algorithms.
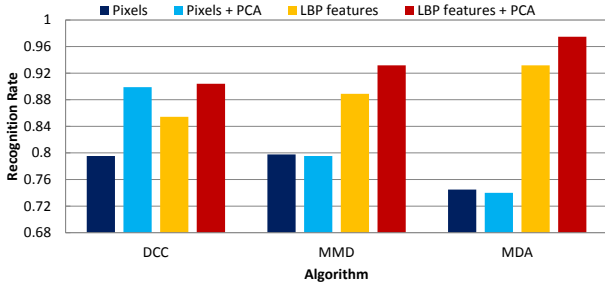

Figure 7. Recognition Rates averaged over three folds for the structure based image set classification algorithms.

to define set similarity. In DCC, the embedding space is also set to 100.

For MMD and MDA, the ratio between Euclidean distance and Geodesic distance is varied from 2.0 to 0.01 in steps of 0.01. Maximum recognition rates were observed at 0.10. The number of connected nearest neighbours for computing geodesic distance is also tested at 14, 12 and 10. The best results were observed at the value of 10. The maximum canonical correlation is used in defining MMD. For MDA, the number of between-class NN local models is set to 5 and the dimension of MDA embedding space is set to 10 as recommended by the authors [19, 18].

6. Results and Discussion

In our experiments, image set based periocular region biometrics has exhibited very good recognition rates for both categories of algorithm. The best results are obtained for the feature vectors of type 4 (LBP followed by PCA). For the sample based algorithms, recognition rates are 94.95%, 96.72%, 96.44%, 97.70% for NN, AHISD, CHISD, and SANP, respectively (Fig. 6). For the structure based algorithms, the best recognition rates are 90.40%, 93.18%, 97.47% for DCC, MMD, and MDA, respectively (Fig. 7). The best identification performance of 97.70% was achieved by SANP. To the best of our knowledge, this is the best identification rate obtained on the MBGC NIR dataset. These results demonstrate the validity of periocular region based human identification using our image set
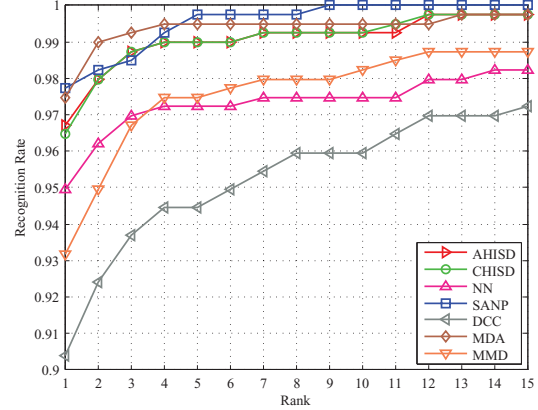

Figure 8. CMC curves for seven image set classification algorithms for feature type 4 (LBP followed by PCA).

approach.

We observe significantly better recognition rates for LBP features as compared to the raw pixel values. One of the reasons may be that LBP filtering reduces the impact of illumination variations and generates an illumination invariant representation [12]. LBP also takes into account the local pixel neighborhoods, therefore, resulting in more discriminative texture patterns. The LBP features also performed better in the case of feature vectors of type 2. This shows that the LBP features may also be employed if high resolution periocular images are not available.

The Cumulative Match Characteristic (CMC) curves for feature type 4 (LBP features followed by PCA) are also plotted for each algorithm in Fig. 8. SANP [6] algorithm performed the best by achieving 100% recognition rate at rank 9. At rank 4 cumulative recognition rates are ≥99.00% for the AHISD, CHISD, MDA, and SANP while ≥97.00% for MMD and NN.

We also performed verification experiments on the same dataset using the same 3 fold approach. Table 1 shows the average verification rates at False Accept Rate (FAR) of 0.001 for feature type 4. It can be observed that although the nearest neighbour algorithm performed well in the identification scenario, it is not robust in the case of verification. Note that it is not our aim to provide an unbiased comparison of these image set classification algorithms. Our main objective is to show the feasibility of image set based clas-

Table 1. Average Identification rates and verification Rates at 0.001 FAR for feature type 4 (LBP followed by PCA)

Method	Identification Rate	Verification Rate
NN	94.95%	55.40 %
AHISD[4]	96.72%	87.65 %
CHISD[4]	96.44%	87.65 %
SANP[6]	97.70%	87.65 %
DCC[9]	90.40%	69.13 %
MMD[19]	93.18%	71.72 %
MDA[18]	97.47 %	86.87 %

Table 2. Execution time in seconds for 11220 image set to image set comparisons using feature type 4 (LBP followed by PCA).

Method	Training Time	Testing Time
NN	N/A	4.11s
AHISD[4]	N/A	10.06s
CHISD[4]	N/A	161.96s
SANP[6]	N/A	> 300s
DCC[9]	4.75s	4.31s
MMD[19]	3.0s	6.12s
MDA[18]	9.31s	5.07s

sification for periocular biometrics.

Table 2 shows the total execution time for matching 132 probe sets with 85 gallery sets using Matlab implementations on a 3.4GHz CPU with 8GB RAM. We observe that SANP, AHISD, and CHISD are computationally more expensive than the structure based techniques. The simple NN technique is the fastest.

7. Conclusion and Future Work

In this paper, we investigated the performance of periocular region biometrics in the framework of image set classification. For this purpose, we performed extensive experimentation with six state-of-the-art image set classification algorithms using four different feature types. Our results demonstrate that the set based approach performs significantly better than the previous single image based techniques. We also observed that the use of LBP features can give better performance than using raw pixel values. This work may be extended by testing the performance of periocular region biometrics on other datasets, especially the MBGC visible spectrum dataset.

8. Acknowledgements

References

[1] Multiple Biometric Grand Challenge (MBGC) dataset http://face.nist.gov/mbgc/.

[2] J. Adams, D. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn. Genetic-based type ii feature extraction for periocular biometric recognition: Less is more. In *ICPR*, pages 205 –208, Aug 2010.

[3] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn. Image understanding for iris biometrics: A survey. *Comput. Vis. Image Underst.*, 110(2):281–307, May 2008.

[4] H. Cevikalp and B. Triggs. Face recognition based on image sets. In *CVPR*, pages 2567 –2573, Jun 2010.

[5] K. Hollingsworth, S. Darnell, P. Miller, D. Woodard, K. Bowyer, and P. Flynn. Human and machine performance

[6] Y. Hu, A. Mian, and R. Owens. Face recognition using sparse approximated nearest points between image sets. *IEEE Transactions on PAMI*, 2012.

[7] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, and C. Suen. Investigating age invariant face recognition based on periocular biometrics. In *IJCB*, pages 1 –7, Oct 2011.

[8] F. Juefei-Xu and M. Savvides. Unconstrained periocular biometric acquisition and recognition using cots ptz camera for uncooperative and non-cooperative subjects. In *Workshop on Applications of Computer Vision (WACV)*, 2012.

[9] T.-K. Kim, J. Kittler, and R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. *IEEE Trans. on PAMI*, 29(6):1005 –1018, Jun 2007.

[10] Y. Lee, R. Micheals, and P. Phillips. Improvements in video-based automated system for iris recognition (vasir). In *Workshop on Motion and Video Computing*, pages 1–8, Dec 2009.

[11] Y. Lee, P. J. Phillips, and R. J. Micheals. An automated video-based system for iris recognition. In *Int. Conf. on Advances in Biometrics*, pages 1160–1169, 2009.

[12] S. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. on PAMI*, 29(4):627 –639, Apr 2007.

[13] P. E. Miller, A. W. Rawls, S. J. Pundlik, and D. L. Woodard. Personal identification using periocular skin texture. In *Proceedings of the ACM Symposium on Applied Computing*, pages 1496–1500, 2010.

[14] U. Park, R. Jillela, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum. *IEEE Trans. on Information Forensics and Security*, 6(1):96 –106, Mar 2011.

[15] U. Park, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *Biometrics: Theory, Applications, and Systems*, volume 1, pages 1–6, Sep 2009.

[16] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR*, volume 1, pages 947 – 954, Jun 2005.

[17] T. Ahonen, A. Hadid and M. Pietikainen. Face Description with Local Binary Patterns: Application to Face Rec. *IEEE Trans. on PAMI*, 28(12):2037–2041, 2006.

[18] R. Wang and X. Chen. Manifold discriminant analysis. In *CVPR*, pages 429 –436, Jun 2009.

[19] R. Wang, S. Shan, X. Chen, and W. Gao. Manifold-manifold distance with application to face recognition based on image set. In *CVPR*, pages 1 –8, Jun 2008.

[20] D. Woodard, S. Pundlik, J. Lyle, and P. Miller. Periocular region appearance cues for biometric identification. In *CVPRW*, volume 1, pages 162 –169, Jun 2010.

[21] D. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *International Conference on Pattern Recognition*, volume 1, pages 201 –204, Aug 2010.

[22] J. Xu, M. Cha, J. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides. Robust local binary pattern feature sets for periocular biometric identification. In *Biometrics: Theory Applications and Systems*, pages 1 –8, Sep 2010.

on periocular biometrics under near-infrared light and visible light. *IEEE Transactions on Information Forensics and Security*, 7(2):588 –601, Apr 2012.