Combining 3D and 2D for less constrained periocular recognition

Conference or Workshop Item

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Combining 3D and 2D for less constrained periocular recognition

Lulu Chen and James Ferryman
Computational Vision Group, School of Systems Engineering
University of Reading, Whiteknights, Reading, RG6 6AY, UK
{l.chen, j.m.ferryman}@reading.ac.uk

Abstract

Periocular recognition has recently become an active topic in biometrics. Typically it uses 2D image data of the periocular region. This paper is the first description of combining 3D shape structure with 2D texture. A simple and effective technique using iterative closest point (ICP) was applied for 3D periocular region matching. It proved its strength for relatively unconstrained eye region capture, and does not require any training. Local binary patterns (LBP) were applied for 2D image based periocular matching. The two modalities were combined at the score-level. This approach was evaluated using the Bosphorus 3D face database, which contains large variations in facial expressions, head poses and occlusions. The rank-1 accuracy achieved from the 3D data (80%) was better than that for 2D (58%), and the best accuracy (83%) was achieved by fusing the two types of data. This suggests that significant improvements to periocular recognition systems could be achieved using the 3D structure information that is now available from small and inexpensive sensors.

1. Introduction

Identifying people using the region around their eyes, known as periocular recognition, has recently become an active research topic in biometrics. Background research has been conducted to prove the importance and strength of using the periocular region for biometric identification tasks [13], in particular for situations where the facial region is largely covered or occluded and long-distance iris capture fails. For these reasons, as a novel biometric trait, periocular has primarily been considered for combination with face [14, 15] and iris recognition [23, 17, 21, 16] to enhance overall recognition accuracy.

Current research on periocular recognition has focussed only on the 2D texture information captured under either visible spectrum or near infrared. 2D images present appearance cues that can be used to distinguish identities. However, in contrast to using a whole face, a single peri-
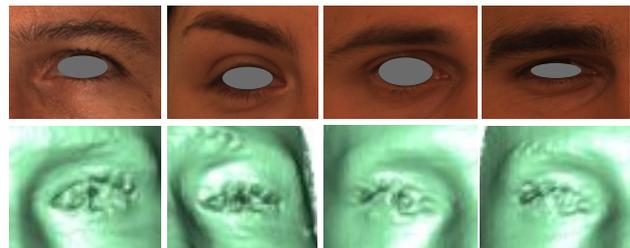

Figure 1. Examples of periocular region data in 3D and 2D [18], top row: colour images, bottom row: rendered 3D images.

ocular region is much smaller, thus the appearance features are limited. In particular, the appearance can be heavily affected by factors including lighting conditions, eye makeup, and natural skin aging. The reported recognition performance from the literature has mainly focussed on ideally captured periocular images [4].

3D information captured from a 3D scanner or a depth sensor holds important cues to a person's facial structure, which can be used for person recognition. The 3D structure of a human face is not affected by facial makeup and skin changes, and 3D capture is more resistant to illumination variation. New depth sensors, such as Microsoft Kinect, have enabled applications for daily life, such as gaming, and natural human-machine interaction with devices like computers and TVs. More recently, Google launched Project Tango [1] that has a depth sensor built into a mobile device, so that 3D information can be captured for real-time tasks like 3D motion tracking and 3D environment reconstruction. Affordable prices and decreases in size mean that a 3D sensor could now be easily integrated into an existing biometric system.

Similar to face recognition in general, periocular recognition must address intra-class challenges, due to illumination variance, facial expression, eye movement, head pose, eye makeup, partial occlusions, etc. and inter-class challenges from similarities between subjects. It is very difficult for 2D alone to cope with these variations, especially when the captured periocular region is small. Therefore, we
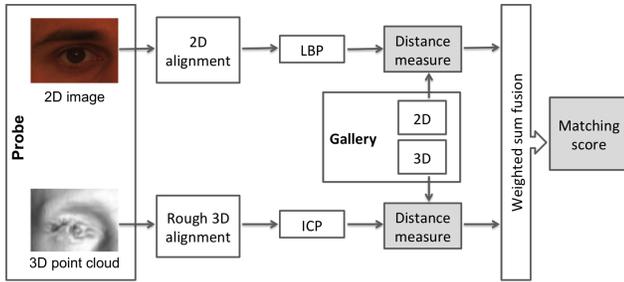
Figure 2. Flow chart of the proposed fusion scheme of 3D and 2D for periocular recognition.

believe combining 3D shape cues, which provide strong 3D structural information of the eye region, and 2D appearance cues, will lead to increased accuracy.

This paper presents the first study that combines 3D global structure and 2D local texture for periocular recognition. The proposed approach does not require training which means that new subjects can be easily added into the system, and it is simple to implement. The approach was evaluated using cropped periocular regions from the Bosphorus 3D face database. The recognition accuracy was calculated for 2D alone, 3D alone, and score-level fusion of both. The 3D was more robust than the 2D against variation in pose and facial expression. The best result was achieved by fusing the two types of data. Figure 2 illustrates the workflow of the proposed fusion scheme of 3D and 2D for periocular recognition.

The rest of the paper is organised as follows. A brief introduction on the literature is presented in Section 1.1. The proposed approaches for 2D, 3D and fusion are described in Section 2, Section 3 and Section 4 respectively. The dataset used, experimental results and evaluation are discussed in Section 5. Finally, conclusions are drawn in Section 6.

1.1. Related work

The periocular region was recently introduced as a novel biometric trait [13], and related research interest has grown quickly since then. Park *et al*. applied local binary patterns (LBP) and SIFT as descriptors for matching the periocular region. Later, 2D image texture features such as LBP, HOG, Gabor and SIFT became popular for this task under visible spectrum [4, 24, 7], near infrared [6, 22] and combined cases [3, 20].

Many of the earlier works have only evaluated their performance on a near ideal periocular dataset, where the subjects face the camera and little variation is present. However, in practical situations the appearance around the eyes will change under factors like illumination, pose, facial expression and occlusion, which make it more challenging to develop a system that is robust for less constrained periocular recognition.

Padole and Proenca [12] first investigated the effect of these factors on recognition performance, and showed that the performance decreased accordingly. Recent works have started incorporating unconstrained periocular images, where the dataset contains this type of variation [8, 9] and recognition is therefore more challenging. Mahalingam and Ricanek presented a multi-scale hierarchical three-patch LBP framework to address the challenges. The performance was evaluated on several challenging datasets that contain occlusions, variation in head pose and closed eyes. Their proposed method outperformed the traditional LBP methods. Xu and Savvides [8] presented discrete transform encoded local binary patterns (DT-LBP) with subspace representation analysis. They conducted several experiments and evaluations on the FRGC v2 database which contains variation in illumination, distance, focus, hair occlusion, motion blur and small degrees of facial expression.

The current state of the art in periocular recognition has only been based on 2D image data. Recognition of the whole face using 3D data has shown its advantages [2] during the past decade. 3D sensing technology has been developed in recent years, and we believe that it is the right time to introduce 3D sensing into periocular recognition.

In this work, the effectiveness of fusing 2D and 3D data for periocular recognition is explored by combining a simple existing approach for 2D (LBP), with another established approach from 3D whole-face recognition (ICP). The two are fused using a simple weighted sum, which is easy to implement. The experimental results showed that using 3D data alone outperforms using 2D, in the presence of a large selection of variations in facial expression and head pose. And more importantly, fusing the two types of data gives a better result than using either one alone.

2. Recognition using 2D texture

The Bosphorus face database contains face data in pairs: a 2D colour image and a 3D point cloud. Figure 1 illustrates a few examples of these pairs from the database, after the periocular region has been cropped. As mentioned in the section above, various texture descriptions have been used for 2D periocular recognition. Because of LBP's power in texture analysis, and its wide use in biometric recognition, in this work the extended circular LBP [10] is applied for matching the 2D images.

2.1. Region of interest and normalisation

To obtain the 2D periocular region, the face images are registered and cropped. An arbitrary neutral face image from the database is selected as the reference face. The faces in both gallery and probe sets are translated, rotated, and scaled to align with the reference face. The transformation is calculated using facial landmarks at the outer corners of the left and right eyes, and the nose tip. The full face

images are then cropped to create two separate periocular regions of equal size (475 × 420 pixels), one for each eye.

2.2. Local Binary Patterns

Local Binary Patterns (LBP) [11] have been widely applied in biometric recognition. The literature describes variations of them used for face recognition, counter-spoofing, and iris recognition, due to their discriminative power for finding fine details on human skin, and their computational efficiency. Thus, LBP is chosen in this work to create a descriptor of the 2D periocular image.

Each periocular image is divided into a grid of $8 \times 8 = 64$ sub images. Circular LBP (CLBP) [10] histograms are calculated for each sub image. The histograms are concatenated to form one long descriptor, to represent the 2D image. To compute the similarities for each probe image against the gallery images, a χ^2-distance measure is used,

$$\chi_D^2(H_p, H_g) = \sum_{i,j} \frac{(H_{p_{i,j}} - H_{g_{i,j}})^2}{H_{p_{i,j}} + H_{g_{i,j}}}, \qquad (1)$$

where H_p is the CLBP histogram descriptor for each probe image and H_g is the equivalent descriptor for each gallery image. i and j represent the i-th bin of the histogram and the j-th sub image.

3. Recognition using 3D shape matching

Due to recent advances in 3D sensing technologies, 3D recognition techniques have been applied to biometrics, such as face recognition and counter-spoofing. 3D information provides the real world geometric structure of the face, which is more resistant to scaling, head poses and illumination changes than 2D images. Also, the 3D shape provides more discriminative features for person recognition.

A common approach for 3D face recognition is to perform rigid face registration on the 3D point cloud. Each probe face is aligned against each of the gallery faces, and the sum of the point distances is calculated as the distance measure. Iterative Closest Point (ICP) has been favoured for this task in 3D face recognition. One of the main challenges in 3D face recognition is large degrees of facial expression. A rigid ICP method has limited ability to handle dramatic non-rigid changes to the face shape. However, the periocular region is only a small region of the face, and the 3D structure around the eye presents less variation than other parts of the face, such as the mouth region. Thus, ICP is suitable for the small periocular region, even under non-rigid expressions.

3.1. Normalisation

To perform ICP robustly against head poses, a rough rigid 3D transformation is performed using facial landmarks. Similar to the normalisation of the 2D images, an arbitrary neutral 3D face from the database is selected as the reference face. The other 3D faces from both gallery and probe are translated and rotated rigidly to match the landmark positions. This achieves a rough alignment of the head pose relative to the neutral reference face. Singular value decomposition (SVD) is applied for finding the optimal rigid transformation in a least-squares sense.

The periocular region is then cropped using the corresponding coordinates from the 2D images, which results in the same cropped region for 2D and 3D for both left and right eyes.
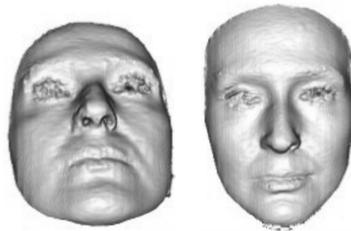

Figure 3. 3D rendered faces: with pitch rotation (left) and the neutral face (right) to which it should be aligned.

3.2. Iterative Closest Point matching

Iterative closest point (ICP) is an algorithm for finding the optimal alignment between two point clouds [5]. It has been used in previous work for 3D shape registration for whole-face recognition.

An initial rough pose is required for the probe point cloud. In this work, this is a rigid transformation calculated from the landmark points in the probe and gallery faces, as described in Section 3.1 above. The periocular 3D point cloud from the probe is then matched against each of the gallery periocular 3D point clouds using ICP, which results in a distance matching score for each one.

The steps for ICP are as follows:

1. For each point in a probe periocular point cloud, the closest point in the gallery periocular point cloud is found by iterating through the points.

2. A rigid transformation is estimated to minimise the distances between all pairs of points, which can be quickly computed using a least-squares solution.

3. The points in the probe periocular point cloud are transformed using the estimated rigid transformation.

The algorithm iterates the steps until finds the optimal rigid transformation for all pairs of points. The distance metric, is then calculated using the sum of the distances from the points in the probe point cloud to the nearest points in the gallery point cloud. The distance measures how well

the two 3D point clouds fit, i.e. the smaller the distance is, the closer the two point clouds match.

4. Fusion

One of the focusses of this work is to evaluate and compare recognition performance of 3D versus 2D data. Also, we would like to see if fusing these two modalities enhances the overall performance. The fusion of 3D and 2D is performed at the matching distance score level. The matching distance scores obtained for each pair of 3D ICP distance and 2D CLBP histogram distance are normalised using median values to make them comparable with each other and reduce the influence of outliers. Then the fusion distance d_f is calculated as a weighted sum:

$$d_f = \alpha * d_{2D} + (1 - \alpha) * d_{3D}, \qquad (2)$$

where α is the fusion weight $0 \leq \alpha \leq 1$, and d_{2D} and d_{3D} are the normalised distance scores for 2D CLBP and 3D ICP respectively.

5. Experimental results and analysis

The performance of the proposed approach was evaluated on the Bosphorus 3D face database [18, 19], which contains a total of 4666 face scans from 105 subjects. Each one has a visible-spectrum image and a 3D point cloud. The capture distance of the subjects is about 1.5 metres, with resolutions of 0.3mm (x), 0.3mm (y) and 0.4mm (depth). The high quality colour images were captured under homogeneous lighting by applying a 1000W halogen lamp in a dark room. Most previous works have used an iris database for testing periocular recognition, that is, most of the images are frontal eye images without variations in pose and facial expression. In comparison, the Bosphorus database contains various head poses and facial expressions, and partial occlusion.

One pair of neutral expression periocular data for each subject was selected for the gallery pool. Thus, the gallery pool consisted of 210 sets of 3D and 2D data, for the left and right eyes of all of the subjects. A total of 7422 of the remaining cropped eye regions were used as the probe.

5.1. 3D versus 2D

Performance was evaluated in $1 : N$ identification mode. Rank-1 recognition accuracy for the ICP-based 3D approach, the LBP-based 2D approach, and the 3D+2D fusion are shown in Table 1. Using 3D structure was over 20% better than 2D texture. The periocular region is much smaller than the whole face, thus 2D LBP texture can provide only limited features for inter-class discrimination. Facial features around the eye, like the eyebrow and eyelid, have proved to be useful for periocular recognition [4]. However, in practice, eye makeup and natural aging can easily change the appearance of these features, which is a challenge to using 2D appearance. On the contrary, the 3D shape of the periocular region is more robust and discriminative than its appearance.

Approach	Rank-1 accuracy
3D	80.48%
2D	57.44%
Fusion 3D+2D $\alpha = 0.25$	83.35%

Table 1. Rank-1 classification results for 3D ICP, 2D LBP and fusion of 3D+2D ($\alpha = 0.25$).

Of all the probes, 3.25% failed using 3D but suceeded using 2D, whereas 27.51% failed using 2D but suceeded using 3D. Figure 6 gives some example for these two scenarios. The left column shows two examples of failed cases using the 2D approach while the middle column shows two examples for failed 3D cases. These demonstrated examples all failed at rank-10 with one data type (2D or 3D), whilst suceeding at rank-1 with the other. For 3D, most failed cases were caused by occlusion by glasses. The glasses obscure the 3D shape of the eye region which significantly decreases recognition performance. In contrast, 2D features are less affected as the glass lenses are transparent. For the 2D approach, however, most failed cases are caused by facial expressions and head rotations.

The power of ICP is to achieve a rigid 3D surface alignment globally across the periocular region. Although under facial expressions, the shape of local regions of the facial surface will change, the overall 3D shape of the person's face cannot be significantly altered. Facial shape changes under expressions are non-rigid, but they mostly affect the bottom half of the face, and the recognition results suggest that using ICP on the periocular region alone is accurate enough, even in the presence of changes in expression.

5.2. 3D + 2D

The recognition results show that combining 3D and 2D using a weighted sum led to a slight increase in accuracy. Figure 4 plots the rank-1 recognition accuracy under different α values. When $\alpha = 0.25$, fusion achieves the best result of 83.35%. Figure 5 illustrates the Cumulative Match Characteristic (CMC) results for 3D, 2D and fusion. It shows that the fusion of global shape and local texture appearance consistently delivers better accuracy than either method alone.

The right column in Figure 6 shows two examples that were correctly classified at rank-1 when combining 2D and 3D, but failed at rank-5 when using 2D or 3D alone. The results indicate that combining 2D and 3D features overcomes the large variations in both facial appearance and 3D shape caused by, for example, intense facial actions and expressions around the eye region. Of all the cases when classifi-
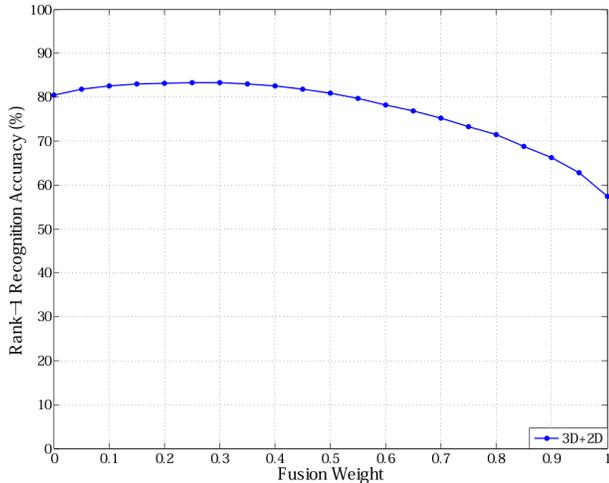
Figure 4. Fusion curves of using different weights. $\alpha = 0$ means 3D only, $\alpha = 1$ means 2D only.
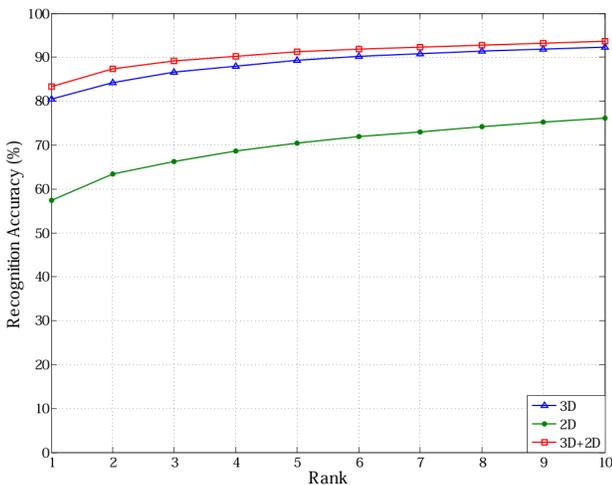

Figure 5. Cumulative Match Characteristic (CMC) results for 3D, 2D and fusion ($\alpha = 0.25$).

cation failed using both 2D and 3D alone, 21.94% became correctly classified when fusing 2D and 3D together.

5.3. Intra-class variations

The Bosphorus 3D dataset [18] contains pose variations, facial expressions, eye wear, and illumination changes. Figure 7 shows some examples of the challenging cases, which are close to real-life scenarios.

The low accuracy of the 2D approach indicates that using a single neutral expression image is insufficient to handle large appearance changes. In other words, with the presence of large variations in facial expression and head pose, LBP texture description fails to interpret these challenging
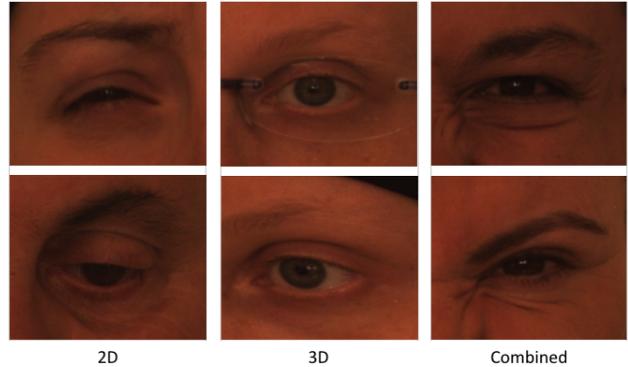

Figure 6. Examples of classification results. Left: failed examples for 2D only. Middle: failed examples for 3D only. Right: examples that were only successful when combining 2D and 3D.

less constrained scenarios, and would require more images, which contain these variations per subject, to be put into the gallery. In comparison, 3D structural data for the periocular region seems to give stronger cues, and be more robust against these variations. The rigid ICP surface alignment captures the global 3D shape, which gave better results.

Table 2 breaks down the recognition results by the classes of variation in the dataset. Rank-1 accuracy is shown for 3D and 2D on each class. Lower face actions (LFAU) and action unit combinations (AUC), such as various mouth movements, have negligible effect on the eye region, so both 3D and 2D gave good performance for these classes that is close to that on neutral expression and pose (N). For upper face actions (UFAU) such as eye closure and eye brow movements, emotional expressions (E) such as disgust and surprise, and different head rotation poses (YR, PR, CR), appearance changes on the eye region are significant. For these classes, 3D maintained reasonable robustness, whereas 2D performance was significantly decreased. Yaw rotations are left and right, pitch rotations are up and down, and cross rotations are a combination of both. When the eye region suffers partial occlusion from hair or glasses, a large portion of the data is omitted, and both 3D and 2D suffer from low accuracy.

6. Conclusion

3D sensors have become available and affordable for consumer use, and have shrunk to a small and convenient size, so they could now be integrated into existing biometric systems. This work presents the first study on combining 3D shape matching and 2D texture description for less constrained periocular recognition. The approach was evaluated on a 3D face database with large variations in facial expression and head pose, to test its suitability for real-life applications. 3D provides geometric structure information, which is more resistant to scaling, head poses and illumina-
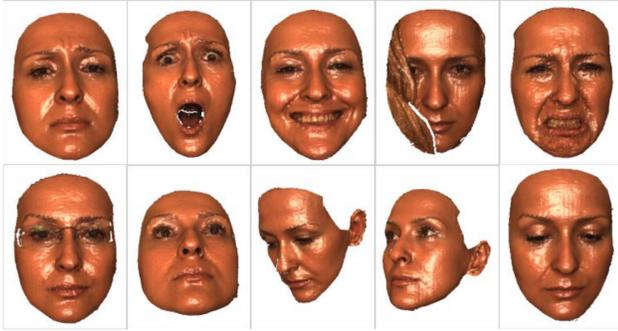
Figure 7. Intra-class variations: examples of facial expressions and head poses from a single subject in the Bosphorus dataset [18].

Facial action class	3D rank-1	2D rank-1
Neutral pose and expression (N)	96.39%	83.25%
Lower Face Action Unit (LFAU)	87.67%	73.79%
Upper Face Action Unit (UFAU)	68.29%	46.99%
Action Unit Combination (AUC)	93.49%	81.95%
Emotional expression (E)	69.09%	57.28%
Yaw Rotation (YR)	69.52%	9.96%
Pitch Rotation (PR)	88.76%	22.25%
Cross Rotation (CR)	59.09%	2.27%
Occlusion (O)	38.56%	45.13%

Table 2. Rank-1 classification results for 3D and 2D against facial expressions and poses.

tion changes. Rank-1 recognition accuracy showed that the 3D significantly outperformed the 2D, which demonstrates the strength of 3D structure over 2D images in overcoming these variations. Moreover, the results suggest that for the small periocular region, 3D shape around the eye region is less affected by large degrees of facial expressions than 2D image appearance. Increased accuracy was consistently achieved by fusing 3D and 2D using a weighted sum, showing that a fusion of the two was better than either alone.

In future work, a 3D deformable model will be considered for dealing with further unconstrained variations that could occur in real-life scenarios.

References

[1] Google Project Tango, last accessed 8 April, 2015, https://www.google.com/atap/project-tango/.

[2] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, 10 2007.

[3] F. Alonso-Fernandez and J. Bigun. Best regions for periocular recognition with NIR and visible images. In *IEEE International Conference on Image Processing*, pages 4987–4991, 2014.

[4] S. Bharadwaj, H. Bhatt, M. Vatsa, and R. Singh. Periocular biometrics: When iris recognition fails. In *BTAS*, pages 1–6, 2010.

[5] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. In *IEEE International Conference on Robotics and Automation*, volume 3, pages 2724–2729, 1991.

[6] K. Hollingsworth, K. Bowyer, and P. Flynn. Identifying useful features for recognition in near-infrared periocular images. In *IEEE BTAS*, pages 1–8, 2010.

[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, 2011.

[8] F. Juefei-Xu and M. Savvides. Subspace-based discrete transform encoded local binary patterns representations for robust periocular matching on NIST's face recognition grand challenge. *IEEE Transactions on Image Processing*, 23(8):3490–3505, Aug 2014.

[9] G. Mahalingam and J. Ricanek, Karl. LBP-based periocular recognition on challenging face datasets. *EURASIP Journal on Image and Video Processing*, (36), 2013.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE TPAMI*, 24(7):971–987, 2002.

[11] T. Ojala, M. PietikÃinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[12] C. Padole and H. Proenca. Periocular recognition: Analysis of performance degradation factors. In *IAPR International Conference on Biometrics (ICB)*, pages 439–445, 2012.

[13] U. Park, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *BTAS*, pages 1–6, 2009.

[14] K. Popplewell, A. Alford, G. Dozier, K. Bryant, J. Kelly, J. Adams, T. Abegaz, K. Purrington, and J. Shelton. A comparison of genetic feature selection and weighting techniques for multi-biometric recognition. In *Proceedings of Annual Southeast Regional Conference*, pages 205–208, 2011.

[15] R. Raghavendra and C. Busch. Robust 2D/3D face mask presentation attack detection scheme by exploring multiple features and comparison score level fusion. In *International Conference on Information Fusion*, pages 1–7, 2014.

[16] K. B. Raja, R. Raghavendra, and C. Busch. Binarized statistical features for improved iris and periocular recognition in visible spectrum. In *International Workshop on Biometrics and Forensics*, pages 1–6, 2014.

[17] G. Santos and E. Hoyle. A fusion approach to unconstrained iris recognition. *Pattern Recognition Letters*, 33(8):984 – 990, 2012.

[18] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus Database for 3D Face Analysis. In *Biometrics and Identity Management*, volume 5372, pages 47–56. 2008.

[19] A. Savran, B. Sankur, and M. T. Bilge. Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units. *Pattern Recognition*, 45(2):767 – 782, 2012.

[20] A. Sharma, S. Verma, M. Vatsa, and R. Singh. On cross spectral periocular recognition. In *IEEE International Conference on Image Processing*, pages 5007–5011, 2014.

[21] C.-W. Tan and A. Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *IEEE Transactions on Image Processing*, 22(10):3751–3765, 2013.

[22] M. Uzair, A. Mahmood, A. Mian, and C. McDonald. Periocular biometric recognition using image sets. In *WACV*, pages 246–251, 2013.

[23] D. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *ICPR*, pages 201–204, 2010.

[24] J. Xu, M. Cha, J. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides. Robust local binary pattern feature sets for periocular biometric identification. In *IEEE BTAS*, pages 1–8, 2010.