BEYOND THE EYE OF THE BEHOLDER: ON A FORENSIC DESCRIPTOR OF THE EYE REGION

C.G. Zeinstra, R.N.J. Veldhuis, L.J. Spreeuwers

University of Twente
Services, Cybersecurity and Safety Group, Faculty of EEMCS
P.O. Box 217, 7500 AE Enschede, The Netherlands

ABSTRACT

The task of forensic facial experts is to assess the likelihood whether a suspect is depicted on crime scene images. They typically (a) use morphological analysis when comparing parts of the facial region, and (b) combine this partial evidence into a final judgment. Facial parts can be considered as soft biometric modalities and in recent years have been studied in the biometric community. In this paper we focus on the region around the eye from a forensic perspective by applying the FISWG feature list of the eye modality. We compare existing work from the soft biometric perspective based on a texture descriptor with our approach.

Index Terms— Soft biometrics, eye region, forensics, FISWG

1. INTRODUCTION

The biometric community traditionally has focused on highly discriminative modalities such as fingerprint and iris. The discriminative property of fingerprints and DNA trace material is utilized in forensic case work for (a) inclusion/exclusion of suspects and (b) assessment of the likelihood whether a suspect is the source of the trace material. However, in practice, trace material might only consist of for example CCTV footage. In that case, forensic facial experts can use morphological analysis on parts of the facial region and combine the outcome of this analysis into a final assessment of the likelihood whether a suspect is depicted on the crime scene images. The Facial Identification Scientific Working Group (FISWG) [1] has published several recommendations for this comparison process, including a one-to-one checklist [2] that summarizes properties of facial parts. Irrespective of the used facial comparison procedure, its core feature is that it combines a multitude of so-called soft biometric modalities instead of one highly discriminating biometric modality. In recent years, soft biometric modalities have been studied extensively in the biometric community.

This paper focuses on the eye region and combines the forensic approach as described in [2] with a texture based approach as found in biometric literature on the periocular region. A formal anatomical definition of the periocular region does not exist. Often the area around the eye (possibly including the eyebrow and the eyeball) is meant. The goal of this paper is to assess (a) the feasibility of FISWG eye features for verification purposes and (b) how their performance relates to existing texture based feature representation. Although real forensic casework typically involves low quality trace material, the feasibility assessment is done on a limited subset of the FRGCv2 dataset. Therefore, this investigation is preliminary and its results should be considered indicative. Annotation of forensic characteristic details can be an eloborate process, but is not unrealistic in a forensic setting.

This paper is organised as follows. In Section 2 we discuss a selection of related work on the periocular region and describe the FISWG eye features. The methodology is discussed in Section 3, where as in Section 4 the data preparation, FISWG descriptor extraction, and experiments are described. Section 5 discusses the results and finally in Section 6 conclusions are drawn.

2. RELATED WORK

Before the emergence of the periocular region as a soft biometric modality, the meticulously detailed eye region model [3] was introduced as a generative model that is ”capable of detailed analysis (...) in terms of the position of the iris, (...) eyelid opening, and the shape, complexity, and texture of the eyelids.” Mainstream interest in the periocular region as a soft biometric modality was sparked by [4]. This paper combines a local (SIFT) and global approach (HOG/LBP texture description on an array of image patches) into a periocular feature set. Tests were conducted on a specially constructed dataset of 30 subjects and approximately 900 images and on a subset of FRGCv2 consisting of 1704 facial images. It was found that manually selected periocular regions that include the eyebrow area and the eyeball give the highest rank-one accuracy rates. Subsequent research has mainly focused on performance under non ideal conditions, alternative texture descriptors and recognition by humans. For example, [5, 6] investigate the performance of uniform LBP (ULBP) and the influence of image quality on the performance, [7]

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

Table 1. Sublist FISWG characteristic eye components and their descriptors. R/L denotes right/left eye. The prefix in the enumeration refers to (D)erived or (A)nnotated Characteristic Descriptors

Component charasteristic	Characteristic Descriptors
Inter-eye distance	(D1) Distance R/L eye
R/L Fissure Opening	(A1) Shape (D2) Angle
R/L Upper Eyelid	(A2) Superior palperal fold (A3) Folds (A4) Epicanthic fold (A5) Lashes
R/L Lower Eyelid	(A6) Lashes (A7) Folds (A8) Inferior palperal fold (A9) Infraorbital furrow
R/L Sclera	(A10) Blood (A11) Defects (D3) Colour
R/L Iris	(A12) Shape (D4) Position, diameter (D5) Colour (A13) Shape pupil (D6) Pupil pos., diameter
R/L Medial canthus	(A14) Shape caruncle (D7) Angle inner eye
R/L Lateral canthus	(D8) Angle outer eye

uses LBP after a (frequency) transformation of the periocular region, and [8] uses the GIST descriptor. Advanced versions of LBP as 3P-LBP, and hierarchical 3P-LBP [9] are also utilized, yielding a rank-one accuracy of 98% on the challenging Notre Dame twins database [10]. The studies on identifying useful recognition features [11] and the performance of human recognition [12] are particularly interesting as they give insight into what clues humans use during their recognition process. In this paper we choose the ULBP texture descriptor as a representative of the texture descriptors that work well under ideal conditions.

The FISWG description of the eye contains an extensive list of characteristic features. In our work we have identified a large sub set of these features. Some of the features are dropped, because they overlap with other features or follow implicitly from other features. The sub set is shown in Table 1. Each characteristic descriptor is either annotated or derived from annotation. We refer to [2] for the complete list.

3. METHODS

Each annotated characteristic descriptor (A1)-(A14) listed in Table 1 is represented by a 2D point cloud. These point clouds use the same coordinate system in which the right and left medial canthi are mapped to (-1, 0) and (1,0), rectifying the face representation. This is advantageous for the calculation of the derived characteristic descriptors since some of them mandate a rectified face representation. Some of the point clouds represent a shape, while other designate noticeable artifacts. Although parametric or more general shape descriptors such as Fourier Descriptors can principally be used for the former case (see for example our work [13] on the eyebrow modality), initial experiments yielded unsatisfactory results. We instead adopt an appearance based approach for all the annotated characteristic descriptors. An example of these appearance based features is shown in Figure 1. Instead of directly superimposing the point clouds of two images, we define a clipping region $[-4, 0] \times [-2, 2]$ for the right eye and $[0, 4] \times [-2, 2]$ for the left eye. By using a 2D binary bin of size 45×45 for each eye region, every annotated point is assigned to a bin. These values are found empirically: the bin size is a trade-off between precision and robustness. In order to determine the influence of the 14 constituent annotated characteristic descriptors the same approach is used.

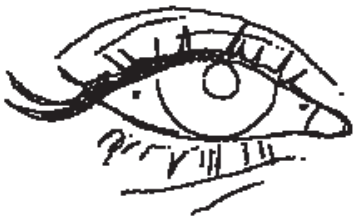


Fig. 1. Example appearance based features

Table 2. Representation of non-appearance features

Non-appearance feature	Representation
(D1) Distance R/L eye	$\mathbb{R}_+$
(D2) Angle eye	$[0, 360]$ (°)
(D3) Colour sclera	$[0, 255]^3 (RGB)$
(D4) Position, diameter iris	$\mathbb{R}^2 \times \mathbb{R}_+$
(D5) Colour Iris	$[0, 255]^3 (RGB)$
(D6) Position, diameter pupil	$\mathbb{R}^2 \times \mathbb{R}_+$
(D7) Angle inner eye	$[0, 360]$ (°)
(D8) Angle outer eye	$[0, 360]$ (°)

The derived characteristic descriptors form the non appearance based features and their representation is shown in Table 2. In order to compare the verification performance with existing periocular literature we use two highly related ULBP approaches. In the first approach we use the original 7×5 grid arranged around the iris as described in [4] for 35 ULBP histograms of 59 bins each. The advantage is that we can compare a basic version of the original descriptor with our ap-

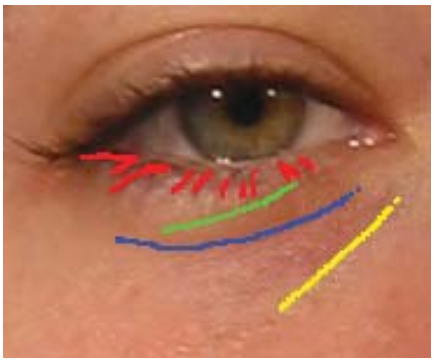


Fig. 2. Example annotation lower eyelid. Annotation from top to bottom are lashes, folds, inferior palperal fold, and infraorbital furrow.

proach. The disadvantage is that it contains the eyebrow area, an area with characteristic features that are not taken into account in our approach. We therefore also use a version that operates on a region of the same size, but shifted downwards with 1.5 bins above and 3.5 bins below the vertical position of the iris center. The original images are gray scaled and rectified based on their medial canthi using bicubic interpolation before ULBP is applied.

4. EXPERIMENTS

4.1. Data preparation and descriptor extraction

For this paper we randomly selected a subset of the FRGCv2 database consisting of 10 persons, 4 images per person from the Spring 2003 session. All images were taken under conditioned illumination and neutral expression. A dedicated annotation tool has been developed in Java. The characteristic descriptors (A1)-(A14) can be annotated in this tool. The user can select and zoom in on parts of the face. An example annotation is shown is Figure 2. The user also manually selects six landmarks on the face: (1) right earlobe connection to the head, (2) right lateral canthus, (3) right medial canthus, (4) left medial canthus, (5) left lateral canthus, and (6) left earlobe connection to the head. Annotation is stored as a collection of points together with the type of characteristic descriptor. The remaining characteristic descriptors (D1)-(D8) are calculated based on the landmarks and annotation. The width of the face is estimated by the intra earlobe landmark distance. The inter-eye distance (D1) is calculated as the distance between the right and left medial canthus in terms of the width of the face. The angle of the eyes (D2) is estimated from the medial and lateral canthi positions. The position and diameter of the iris (D4) are calculated by minimizing the geometric distance [14] to the annotation (A12), the estimation of (D6) is similarly based on (A13). The iris colour (D5) is determined by averaging the colours of the pixels that lie (a) in the fissure opening (A1), (b) in the estimated iris (D4), and (c) outside the estimated pupil (D6). An additional automatic post processing step removes bright artifacts in the iris caused by studio lighting. The color of the sclera is also determined by averaging the colours of pixels that lie (a) in the fissure opening (A1), (b) outside the estimated iris (D5), and (c) outside the caruncle (A14). The inner (D7) angle is estimated by the following procedure: the fissure shape is partitioned into two sets that lie above and below the line segment between the medial and lateral canthus. From those two sets the points that lie outside the proximity of the medial canthus (radius of 10% of the eye width) are removed. The remaining two sets are least squares linearly interpolated and the resulting slopes determine the inner angle. The estimation for the outer angle (D8) is similar.

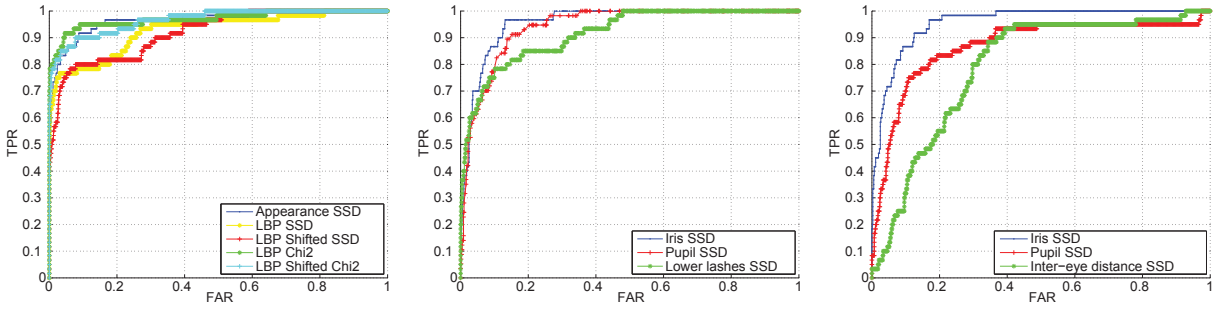
4.2. Experiments

We define three experiments. In Experiment 1 we assess the verification performance of the appearance based features as a group and compare that with the texture based ULBP approaches. In this experiment we use the sum of squared difference and the χ^2 score function. The latter score function is often employed in the LBP case, and is equivalent to the sum of squares differences score function in the (binary) appearance based features approach. In Experiment 2 we determine the verification performance of the appearance based features separately. In Experiment 3 we focus on the verification performance of the non-appearance based features. In Experiment 2 and 3 we use the sum of squares score function. All experiments are conducted on the right and left eye. All presented results are fused on score level by using min-max scaling and sum fusion.

5. RESULTS

In Experiment 1, we compare the verification performance of the appearance approach versus the ULBP and shifted ULBP measured in terms of AUC, resulting in resp. 0.966, 0.926, and 0.919 in the case of the sum of squares score function. We notice that the appearance based method seems to perform slightly better than the texture based methods. However, this difference disappears when the three approaches use the χ^2 score function. In that case the AUC's are resp. 0.966, 0.970, and 0.966. In both cases one might have expected a lower AUC value for the shifted ULBP relative to the AUC value of the original UBLP as the former contains "less" information than the latter one. However, the measured differences seem insignificant. The ROC curves of Experiment 1 are shown in Figure 3a.

The results of Experiment 2 (separate appearance based features) are shown in Table 3. Apart from two trivial outcomes (0.500 in the cases of (A4) Epicanthic fold and (A10) Sclera Blood, caused by absence of annotation) the values of the



(a) Appearance vs LBP, SSD and χ^2 (b) Three best appearance, SSD (c) Three best non appearance, SSD

Fig. 3. Selection of performances of Experiments 1, 2, and 3. The SSD (resp. χ^2) refers to the sum of squared differences (resp. χ^2) score

Table 3. Performance appearance features in terms of AUC

Appearance feature	**AUC**
(A1) Fissure Shape	0.851
(A2) Superior palperal fold	0.820
(A3) Upper Folds	0.758
(A4) Epicanthic fold	0.500
(A5) Upper Lashes	0.808
(A6) Lower Lashes	0.915
(A7) Lower Folds	0.754
(A8) Inferior palperal fold	0.882
(A9) Infraorbital furrow	0.824
(A10) Sclera Blood	0.500
(A11) Sclera Defects	0.787
(A12) Shape Iris	0.957
(A13) Shape Pupil	0.941
(A14) Shape Caruncle	0.762

AUCs vary between 0.754 and 0.957. Surprisingly, the shapes of the iris and the pupil yield the top two AUC values. The shapes reveal the position and size of these modalities and apparently represent a discriminating property. Although every (non-)appearance based feature is easily changed, especially the iris and pupil positions can change instantaneously just by gazing away from the camera. Also the lower lashes, in comparison to the upper lashes, are performing well. This difference might be explained by the fact that upper lashes (a) tend to cover the whole upper fissure opening and (b) are quite dense in their distribution. In contrast, lower lashes when traversing from the medial to the lateral canthus (a) often start around the projection of the iris on the lower fissure opening and (b) exhibit a more sparse distribution. The latter property implies that the lower lashes have a potential to be more ”unique”. These observations are also illustrated by the annotation example in Figure 2. A similar difference can be

Table 4. Performance non-appearance features in terms of AUC

Non-appearance feature	**AUC**
(D1) Distance R/L eye	0.789
(D2) Angle eye	0.700
(D3) Colour sclera	0.726
(D4) Position, diameter iris	0.956
(D5) Colour Iris	0.714
(D6) Position, diameter pupil	0.868
(D7) Angle inner eye	0.773
(D8) Angle outer eye	0.623

observed between the two outlines of the eyelids (Superior palperal and Inferior palperal folds): again, the lower part seems more ”unique”. Apart from the shape of the fissure opening, the remaining four appearance based features AUC fall below 0.800. The study presented in [11] indicates that humans found the eyelashes, tear duct (caruncle), eye shape (fissure opening) and the eyelids most helpful in their identity decision making process while using near-infrared images. This result is clearly reproduced in this study, with one noticeable exception: the caruncle. This might be explained by the difference in the representation of this modality. In our approach only the shape drawn on a limited resolution image exhibiting reflection artifacts (see Figure 2) is taken into account, whereas in [11] the use of the near-infrared spectrum and high resolution demarcates the caruncle very well. In Figure 3b ROC curves of the three best performing appearance based features (iris, pupil, and lower lashes) are shown.

Finally, the verification performance of the non-appearance based features are measured in Experiment 3 and are shown in Table 4. The top two performing modalities are the iris and pupil positions and confirm the performance of their appearance based counterparts. It also indicates that the implicit location and size information contained in annotation

data yield this performance, while other implicit information like fissure opening contributes less. The next best performing feature is the inter-eye distance. Although a very simple measure, it performs relatively well. This is caused by the fact that the localization of the medial canthi is very robust. However, its value expressed as a relative measure can be hampered by approximate localization of the earlobe positions. The inner angle is much more stable than the outer angle, as the localization of the lateral canthi is less robust. This also explains the lower AUC value for the angle of the eyes. Finally, the iris and sclera colours are not very convincing non-appearance features. Overall, the verification performance of the non-appearance features is generally inferior to the appearance based features. ROC curves of the three best performing non-appearance based features are shown in Figure 3c.

6. CONCLUSION

In this paper, we have studied the feasibility of FISWG eye features for verification purposes and how their performance relates to a representative of a texture based feature representation. We find that some of the FISWG features work well (iris, pupil position, either appearance of non-appearance based, lashes, fissure shape), while others are less convincing. Especially the non-appearance features that measure an angle or colour do not perform well. This is in line with a FISWG recommendation "(photo-)anthropometry has limited discriminating power". Using semantically important information such as the FISWG feature set can yield verification performances comparable to texture based methods. Finally, we are very aware that this study has been conducted on a limited subset, so its results should considered indicative.

REFERENCES

[1] "Fiswg website," `http://www.fiswg.org`, Accessed: 2014-04-22.

[2] "Fiswg facial image comparison feature list for morphological analysis," `https://www.fiswg.org/document/viewDocument?id=50`, Accessed: 2014-10-24.

[3] T. Moriyama, T. Kanade, Jing Xiao, and J.F. Cohn, "Meticulously detailed eye region model and its application to analysis of facial images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 5, pp. 738–752, May 2006.

[4] Unsang Park, R. Jillela, A. Ross, and A.K. Jain, "Periocular biometrics in the visible spectrum," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 1, pp. 96–106, March 2011.

[5] Philip E. Miller, Allen W. Rawls, Shrinivas J. Pundlik, and Damon L. Woodard, "Personal identification using periocular skin texture," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, New York, NY, USA, 2010, SAC '10, pp. 1496–1500, ACM.

[6] P.E. Miller, J.R. Lyle, S.J. Pundlik, and D.L. Woodard, "Performance evaluation of local appearance based periocular recognition," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, Sept 2010, pp. 1–6.

[7] Juefei Xu, M. Cha, J.L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust local binary pattern feature sets for periocular biometric identification," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, Sept 2010, pp. 1–8.

[8] S. Bharadwaj, H.S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, Sept 2010, pp. 1–6.

[9] Gayathri Mahalingam and Karl Ricanek, "Lbp-based periocular recognition on challenging face datasets," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 36, 2013.

[10] P.J. Phillips, P.J. Flynn, K.W. Bowyer, R.W.V. Bruegge, P.J. Grother, G.W. Quinn, and M. Pruitt, "Distinguishing identical twins by face recognition," in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, March 2011, pp. 185–192.

[11] K. Hollingsworth, K.W. Bowyer, and P.J. Flynn, "Identifying useful features for recognition in near-infrared periocular images," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, Sept 2010, pp. 1–8.

[12] K.P. Hollingsworth, S.S. Darnell, P.E. Miller, D.L. Woodard, K.W. Bowyer, and P.J. Flynn, "Human and machine performance on periocular biometrics under near-infrared light and visible light," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 588–601, April 2012.

[13] C.G. Zeinstra, R.N.J. Veldhuis, and L.J. Spreeuwers, "Towards the automation of forensic facial individualisation: Comparing forensic to non-forensic eyebrow features," in *Proceedings of the 35th WIC symposium on Information Theory in the Benelux and The 4th Joint WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux*, 2014, pp. 73–80.

[14] Walter Gander, Gene H. Golub, and Rolf Strebel, "Least-squares fitting of circles and ellipses," *BIT Numerical Mathematics*, vol. 34, no. 4, pp. 558–578, 1994.