ON IDENTIFICATION FROM PERIOCULAR REGION UTILIZING SIFT AND SURF

[1]*Şamil Karahan,* [1]*Adil Karaöz,* [1]*Ömer Faruk Özdemir,* [1]*Ahmet Gökhan Gül,* [2]*Umut Uludağ*

[1]Department of Computer Engineering, Gebze Institute of Technology, 41400, Gebze, Kocaeli, Turkey
[2]TUBITAK BILGEM, 41470, Gebze, Kocaeli, Turkey
{ samilkarahan, adilkaraoz, farukozdemir24, ahmetgokhangul }@gmail.com
umut.uludag@tubitak.gov.tr

ABSTRACT

We concentrate on utilization of facial periocular region for biometric identification. Although this region has superior discriminative characteristics, as compared to mouth and nose, it has not been frequently used as an independent modality for personal identification. We employ a feature-based representation, where the associated periocular image is divided into left and right sides, and descriptor vectors are extracted from these using popular feature extraction algorithms SIFT, SURF, BRISK, ORB, and LBP. We also concatenate descriptor vectors. Utilizing FLANN and Brute Force matchers, we report recognition rates and ROC. For the periocular region image data, obtained from widely used FERET database consisting of 865 subjects, we obtain Rank-1 recognition rate of 96.8% for full frontal and different facial expressions in same session cases. We include a summary of existing methods, and show that the proposed method produces lower/comparable error rates with respect to the current state of the art.

Index Terms— Face, Periocular Region, SIFT, SURF, BRISK, ORB, LBP, FLANN, Brute-Force Matcher, FERET, Identification.

1. INTRODUCTION

In many modern applications, face recognition has been the most popular biometric modality for human identification because of (i) ease of acquisition, (ii) omnipresence of sensors (e.g. security/ATM cameras), (iii) cost benefits, and (iv) ability to get the biometric signal without subject cooperation. Despite this popularity, performance of the facial identification methods generally decline when dealing with non-ideal scenarios such as occlusions, non-uniform illumination, expression changes, pose variations, facial appearance changes (presence of beard, glasses, make-up etc.). These problems can be solved to a large extent, either by improving existing algorithms, using more training data, using multimodal biometric systems or exploring a new method to improve existing identification performance [1].

In this work, we have utilized the discriminative potential of periocular region of human faces. This scenario arises when the task is to identify subjects wearing a mask to obscure their faces, or, due to occlusion (by a hat, scarf etc.) or shadows, only that region is available for identification. Performances of typical face recognition algorithms decrease greatly in this scenario. Our goal is to find a discriminative and robust feature set that can be extracted from only that region surrounding eyes, and use this set to perform personal identification, independent of the other facial regions.

The cited periocular region is a small neighborhood around the eyes including eyebrows. Recently, this region has been a popular test-bed, to solve identification problems of occluded face images. The current approaches show that identification can be performed solely based on this periocular region, up to an (expected) performance upper limit.

The periocular region features can be classified as (i) level one-features that include eye lids, eye folds, eye brows and eye corners; or (ii) level-two features that include wrinkles around eyes, and skin pores [1]. In this paper, we extract important keypoints using SURF [2], which uses a Hessian matrix-based metric for the detector, from both level features. For these keypoints, SIFT [3], SURF, BRISK [4] and ORB [5] are used to obtain descriptor vectors. Further, the vectors obtained from SURF and SIFT are concatenated to obtain more details around the keypoints. To observe the effects of LBP [6] method on the periocular region, we split the input images into cells and extract the descriptor vectors pertaining to LBP and SURF from each cell. The final descriptor vector is obtained by concatenating the vectors from each cell.

The eyes are located on the face image symmetrically, so descriptors of some keypoints like eye corners can come close to each other. To overcome this effect, we split the input images into left and right sides and the distance between train and query descriptor vectors is calculated considering these sides using FLANN [7] and Brute-Force feature matching algorithms. The method was evaluated for four different normalization methods during matching phase to determine the best. The algorithm has been tested using grayscale images extracted from FERET dataset

Approach	Feature Extraction Method	Dataset / Subject Number / Image Count	Best Accuracy
Miller et al. [9]	ULBP	FRGC / 410 subjects / 1230 images FERET / 54 subjects / 162 images	FRGC: 89.1% FERET: 74.1%
Adams et al. [10]	LBP + GEFE	FRGC / 410 subjects / 1230 images FERET / 54 subjects / 162 images	FRGC: 92.2% FERET: 85.1%
Woodard et al. [1]	RG color histogram, LBP	FRGC / 410 subjects / 4100 images MBGC /115 subjects / 1700 images	FRGC: 91% MBGC: 87%
Park et al. [8]	HOG, LBP, SIFT	DBI / 30 subjects / 120 images FRGC / 568 subjects / 3408 images	81.6%
Our Method	SURF + SIFT, LBP + SURF, SIFT, SURF, ORB, BRISK	FERET_All / 865 subjects / 2380 images FERET_fa / 153 subjects / 391 images FERET_fb / 150 subjects / 384 images	All: 96.8% fa: 67.3% fb: 63.0%

Table 1. Comparison of state art methods and the proposed method.

and the experimental results are shown in relevant section. Our algorithm performs well in case of images from *same session with facial expression*. Most of the existing research on the periocular biometrics has mostly used dense features which are extracted from the periocular region. The main problem of this method is that it can work well with well-aligned, non-scaled and non-occluded images. The proposed method uses a sparse feature set which is obtained by using scale invariant feature extraction methods. We think that this method is more appropriate for real world scenarios.

We briefly discuss previous work on the periocular region-based personal identification in Section 2. We then describe our data sets in Section 3, give detailed description about each step of the method in Section 4 and present experimental results in Section 5. The conclusions and pointers for future work are given in Section 6.

2. PREVIOUS WORKS

Table 1 shows a summary of previous works in this field, along with performance of our method, detailed in Section 5. Comprehensive applicability of periocular region has been studied by Park et al. [8]. They have detected iris in the visible spectrum image and extracted global and local information from the periocular region using LBP, GO (gradient orientation) and SIFT operators. The Euclidean distance has been used to calculate the matching score of feature descriptor results. The method has been tested using FRGC (Face Recognition Grand Challenge) database, which contains 1704 images and 568 subjects, and creating a dataset, which contains 120 images and 30 subjects. As shown in Table 1, experimental results show a rank-one recognition accuracy of 81.6%.

Some researchers have investigated the use of Local Binary Pattern (LBP) features [6] to represent the texture of the periocular region in a dense feature vector. Miller et al. [9] used city block distance to measure of similarity of the features description. Verification and identification experiments involving 464 subjects were performed on datasets constructed from the FRGC and FERET datasets. Nearly 90% recognition rate has been reported. Their work was extended by Adams et al. [10] using genetic algorithms to select the optimal subset of LBP features and they observed better recognition results. In another work using LBP, Woodard et al. [1] used color histograms to represent periocular region and measured similarity with Bhattacharya coefficient. They have obtained the Rank-1 recognition rate of 91% with FRGC database.

Woodard et al. [11] proposed periocular biometrics as an alternative to iris recognition if the iris images were captured at a distance. The features were extracted from the periocular region using CLBP (Circular LBP) and GIST, which is a global descriptor that describes holistic spatial information of an image. Their observation showed that periocular biometrics can be a good alternative when iris recognition is not feasible, provided it is captured from an optimal distance. Furthermore, for a study analyzing occlusion effects on Yale and AR databases, please see [12].

3. UTILIZED DATA SET

In this work, we use FERET [13] dataset whose photographs were taken from 1993 to 1997. It consists of frontal, left and right image profiles. In order to obtain periocular regions, we use only frontal images whose letter codes are fa and fb in image name.

Frontal images' pose angle is 0 (zero) according to FERET ground truths which have iris, nose, mouth positions. According to FERET documentation;

- fa: indicates a regular frontal image.
- fb: indicates an alternative facial expression frontal image.

We have three dataset scenarios (cf. Table 1, last line). First of them is using all of fa and fb labeled images. In this scenario we limit the number of pictures per subject, as 8. In the others, only fa and only fb labeled images are used to observe the behavior of our work in *different session* and *facial expression* cases.

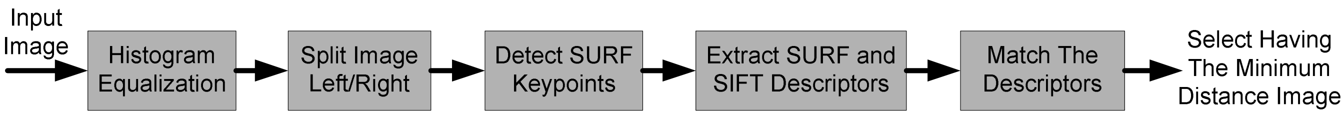

Fig. 1. An overview of our feature extraction and matching system.

4. DETAILS OF THE METHOD

In this section, we describe the steps of our proposed method, which are summarized in Figure 1.

4.1. Histogram Equalization

In the dataset, some images are brighter or darker than others. To solve this problem, that can affect the points to calculate the features, firstly, histogram equalization is applied to images, by adjusting image intensities to enhance contrast. This method usually increases the global contrast of images, especially when the usable data of the image is represented by close contrast values.

4.2. Image Rotation and Split

The extracted features from a specific right periocular region can match with those from left region of other images because of the symmetry present in face images. For this reason, we split periocular images, parallel to the y axis by using position of the irises which are obtained from ground truth files of the dataset. Furthermore, we rotate each image to align the position of irises horizontally. Finally, we drew rectangular boundaries, which were defined by using a certain percentage of Euclidean distance between irises of each side. Fig. 2 shows an example image.

Fig. 2. Input image (left) and processed image (right).

4.3. Keypoint Extraction

Detecting keypoints is the most important problem for scale invariant methods. We use the left and right images separately using SURF as the feature detector. We use it rather than the alternative SIFT detector because SURF is several times faster than SIFT. Further, our experiments show that if we use SIFT algorithm as the detector, the performance decreases.

SURF is a keypoint detector and descriptor which is fast, & scale and rotation invariant. Its algorithm finds key point locations and it outputs a descriptor vector as well.

The selection strategy of keypoints is based on determinant of Hessian blob detector and an integer approximation. This gives a computational time advantage as it only depends on three integer operations.

For the descriptor vector, first step is defining a square area around the keypoint. This region's size depends on diameter of hessian blob. The second step is defining horizontal and vertical weight of Haar wavelet responses. Rotation invariant part of SURF is calculating vertical and horizontal Haar wavelet responses' weights depending on orientation of keypoint's gradient. The last step is the determination of the descriptor vector size and its values. In the original SURF study, using 4×4 sub-regions and descriptor vectors of size 64 are suggested. 4×4 sub-region is found to be the best threshold by experiments, but 128 can be used for size of descriptor vector.

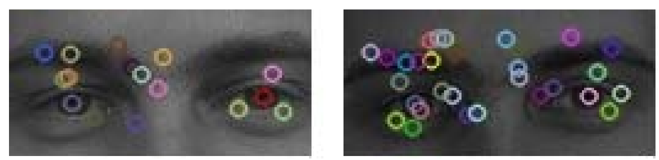

Fig. 3. Detected SURF keypoints of two different images from eye regions of the same subject.

SIFT is similar to SURF. There are several differences on keypoint detector and descriptor parts. In our work, we use SURF keypoint detector so that we only focus on SIFT descriptor part. For the descriptor part, firstly we have to define a good orientation of keypoint and correct with an affine transformation. After that, we compute each pixel's gradient magnitude and orientation values.

The algorithm calculates the histogram by dividing focus area into 4×4 cells and representing the gradient of each pixel with 8 bins so a descriptor vector which has 128 elements is obtained. After the keypoint detection, we compute descriptor vector from keypoints using SIFT and SURF too. SIFT and SURF computes descriptor separately and save descriptor to their associated vectors. Afterwards, we concatenate these vectors side by side. SIFT creates a descriptor vector of 128×N size and SURF creates a descriptor vector of 64×N size. N is equal to the number of keypoints. Fig. 3 shows an example, where identical subjects's two different pictures are overlaid with keypoints detected by SURF.

4.4. Feature Matching

In this section, we discuss our matching strategy. Two different matching algorithms, namely, the Brute-force matcher (basic or naive descriptor matcher) and FLANN

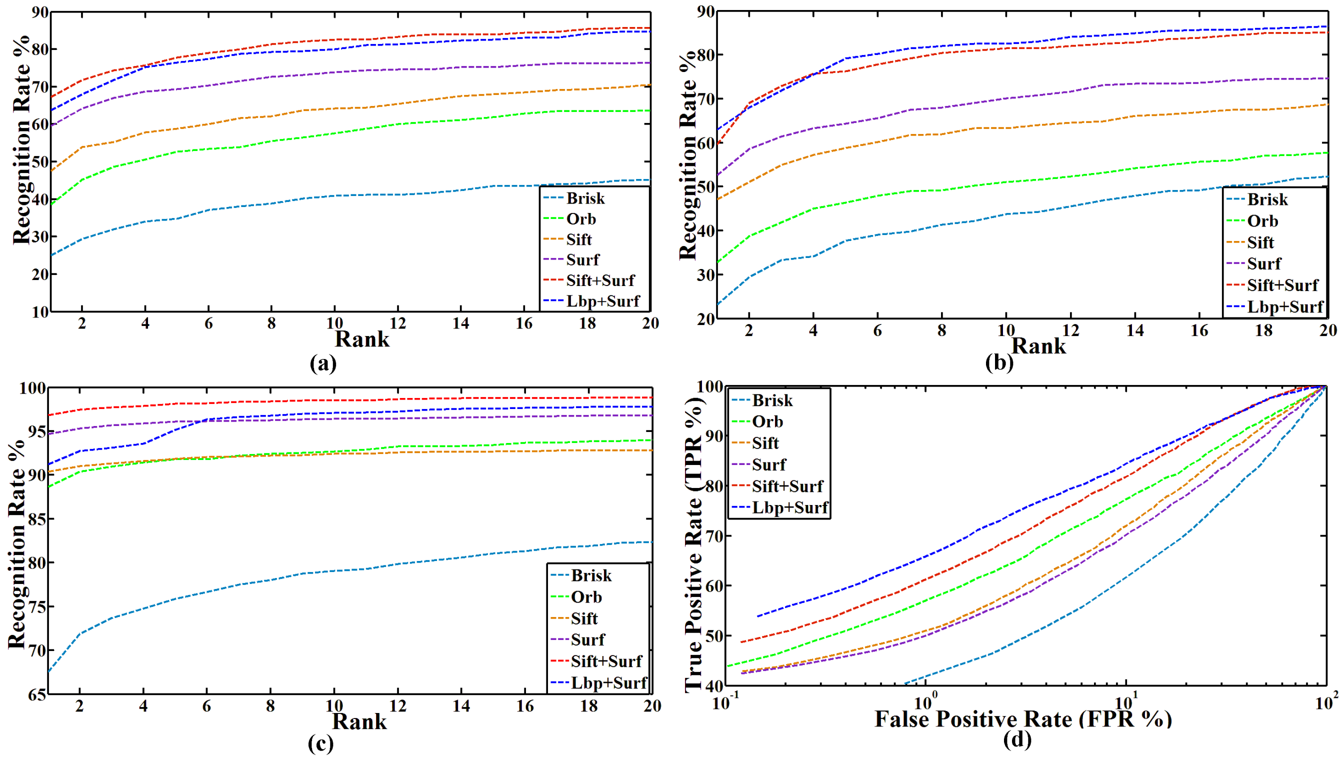

Fig. 4. Performance results of the proposed method: (a) Rank results of fa, (b) Rank results of fb, (c) Rank results of the whole dataset, (d) ROC result of the whole dataset.

matcher, which have been implemented in OpenCV [14], are tested in our studies. After comparing performances on matching algorithms, BF-Matcher outperforms by using L2_NORM for SIFT + SURF, HAMMING_NORM for LBP + SURF and HAMMING2_NORM for ORB. It takes the descriptor of one feature in first set and is matched with all other features in second set using the norm. And the closest one is returned.

4.5. Local Binary Pattern and SURF

In the Local Binary Pattern [6] approach for texture classification, the occurrences of the LBP codes in an image are collected into a histogram. State of the art methods mentioned in Section 2 above, use LBP feature extraction methods for biometric identification from periocular region. In this work, our goal is to show the identification performance by using a combination of SURF and SIFT. Besides, identification performance of combination of LBP and SURF was tested as well. For this operation, we firstly split input image into 64 equal cells to extract descriptor vectors of LBP and SURF from each of them. Finally, the LBP feature vector, which was extracted from each window, and the SURF feature vector, which was extracted from around center point of each cell, was combined to represent the cell descriptors.

5. EXPERIMENTAL RESULTS

We ran our algorithm on three different parts of the FERET dataset’s periocular regions. Number of subjects/images is given in Table 1 above. Park et al. [8] divided periocular region two main parts (left and right). We used the same methodology. Our baseline algorithm is SIFT and SURF. In order to show the performance of the other feature extraction algorithms on periocular regions, we tested the feature extraction methods such as ORB, BRISK, SIFT, SURF, and SURF fusion with SIFT and LBP.

At first experiment, we use whole database has regular face images, different expressions, different session regular and different facial expression images. We observed that our best result is about 96.8% in Rank-1 accuracy. The results of rank score distributions are displayed in Fig. 4 (c). Depicting Linear GAR and logarithmic FAR, ROC result is given in Fig. 4 (d).

Then, we tested our algorithms on different sessions and different expressions so we build our dataset according to session time and expression label. As a result of these images’ illuminations, facial aging and facial expressions are changed. Our accuracy is about 63%; this information is displayed in Fig. 4 (a).

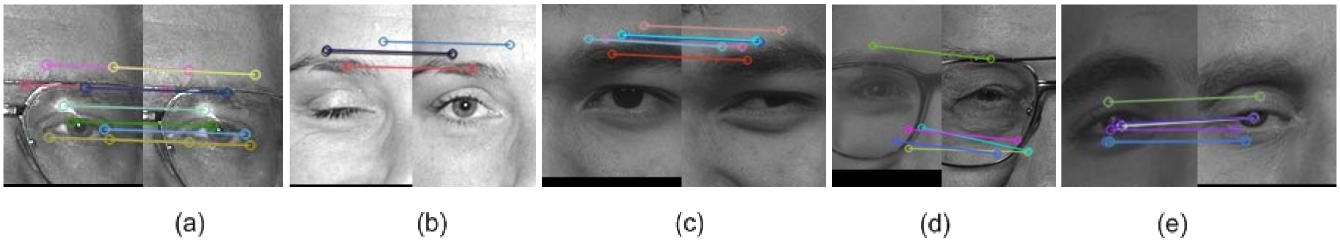

Fig. 5. Example results of the proposed method. (a) TM (true match) with the presence of the glasses, (b) TM in case of the closed eyes, (c) TM in case of the different gaze direction, (d) FM (false match) in presence of the glasses, (e) FM in case of different illumination.

Thirdly, we pick different session frontal faces images to detect temporal facial aging change and illumination changes. We run our algorithm with this specific dataset and obtained the Rank-1 recognition rate of 68%. Associated results can be seen in Fig. 4 (b). We can claim that illumination changes and aging effects are more dominant than different facial expressions. Fig 5 shows several True Match (TM) and False Match (FM) outputs.

Our OpenCV implementation runs on an i5 3.4 GHz CPU computer with 8GB RAM, and execution time for 1 picture matching among 2380 pictures took an average of 3.02 seconds. Our accuracy is better than/comparable with current state of the art, which generally uses small portions of FERET database. Furthermore, in our dataset, average image count of each subject is fewer than other datasets such as FRGC & MBGC.

6. CONCLUSION AND FUTURE WORK

In this paper, we investigated the utilization of various feature extractors for biometric identification from facial periocular region. The experiments presented in this paper measure the performance of the identification on periocular region image data originating from 865 subjects of FERET database. In our experiments, combination of SURF and SIFT gave better performance than using them individually. Moreover, we performed extensive experimentation with four state-of-the-art feature extraction methods which cannot have scale invariant characteristic. For future work, we plan to extend the testing to other datasets and utilizing other machine learning algorithms. Also, different feature extraction methods could be used to handle variation on different sessions to decrease identification error.

7. ACKNOWLEDGEMENT

This work was financially supported by ARDIC Technology (www.ardictech.com)

REFERENCES

[1] D. Woodard, S. Pundlik, J. Lyle, and P. Miller, "Periocular region appearance cues for biometric identification", in *Computer Vision and Pattern Recognition Workshops*, San Francisco, USA , 2010, pp. 162-169.

[2] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features", in *European Conf. Computer Vision*, Graz, Austria, 2006, pp. 404-417.

[3] D. G. Lowe, "Distinctive image features from scale-invariant key points", *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91-110, November 2004.

[4] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints", *Int. Conf. Computer Vision*, Barcelona, Spain, 2011, pp. 2548-2555.

[5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF", *Int. Conf. Computer Vision*, Barcelona, Spain, 2011, pp. 2564-2571.

[6] T. Ahonen, H. Abdenour, and M. Pietikäinen, "Face recognition with local binary patterns", in *European Conf. Computer Vision*, Berlin, Germany, 2004, pp. 469-481.

[7] M. Muja, and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration", in *Int. Conf. Computer Vision Theory and Application (VISSAPP)*, Lisboa, Portugal, 2009, pp. 331-340.

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

[9] P. E. Miller, A. W. Rawls, S. J. Pundlik, and D. L. Woodard, "Personal identification using periocular skin texture", *Proc. ACM Symp. Applied Computing*, Sierre, Switzerland, 2010, pp. 1496-1500.

[10] 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", *Int. Conf. Pattern Recognition*, Istanbul, Turkey, 2010, pp. 205-208.

[11] D. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross, "On the fusion of periocular and iris biometrics in non-ideal imagery", *Int. Conf. Pattern Recognition*, Istanbul, Turkey, 2010, pp. 201-204.

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust face recognition via sparse representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, no. 2, December 2008, pp. 210-227.

[13] P. J. Phillips, H. Wechsle, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms", *Image and Vision Computing*, vol. 16, no. 5, April 1998, pp. 295-306.

[14] OpenCV, http://docs.opencv.org/modules/refman.html.