Human Facial Feature Matching based on Motion-Smoothness Constraint

Liguo Dong¹*, Wenchao Xu¹ and Xianxian Zeng²

¹ South China Institute Of Software Engineering, Guangzhou University, Guangzhou, Guangdong, 510990, China
² Guangdong University of Technology, Guangzhou, Guangdong, 510006, China
* E-mail: 1062491936@qq.com

Abstract. Face recognition is a practical and important problem. In our life, face recognition problems exist everywhere. How to identify the face in the image accurately and quickly is an important and urgent topic. In this paper, an efficient and accurate face recognition method based on keypoints feature matching is proposed to extract the keypoints features of the face image, and then the nearest neighbor method is used to obtain a large number of keypoints rough matching. Then the correct matching of the keypoints is determined by the motion smoothness constraint of the original keypoints matching. Finally, the matching of the face is determined according to the number of keypoints matching between the two input face images. Some experiments show that this method is suitable for face recognition with expression change and pose change.

1. Introduction
Human face recognition and face identify are a very import issue and a lone-standing issue. Many scholars and companies have devoted a lot of energy and financial resources to making recognition performance better and faster. In the past, Turk et. al. proposed the Eigenface method [1], using Principal Component Analysis (PCA) to obtain the appropriate vector to describe the input face image for face recognition. Later, the classic algorithm of face recognition, namely Fisherface [2], was proposed by Belhumeur et. al. Fisherface was based on linear discriminant analysis, and had similarities with principal component analysis, so it was not robust to illumination. Later, Ahonen et. al. proposed Face recognition with local binary patterns (LBP) [3], which was the extraction of local features. Although it was not sensitive to light, it still was not robust for pose and expression.

With the increase of data volume and the development of parallel computing hardware, such as GPU, some deep learning methods [4][5] for face recognition were proposed. These methods used a mount of data to train a convolutional network model. And the trained model can be correctly identified in a lot of cases (small pose changes, light changes and expression changes). However, deep learning method requires a large of training data. If the training data are not enough, the model will overfit to training data.

The classical image methods and deep learning methods did not obtain good performance, and some scholars proposed to use local keypoint feature matching to replace the traditional face recognition method. In recent years, Li et. al. proposed Pore-SIFT (PSIFT) [6], which was a variant of Scale-invariant feature transform (SIFT) [7]. Li et. al. analyzed the characteristics of the human skin, and reform the detector of SIFT to get more keypoints. After getting keypoints, PSIFT used the
descriptor of SIFT to get the keypoints features, and then the nearest neighbor proportional method was used as the matching method for matching keypoints features. This method was robust to light, pose and expression changing. Furthermore, PSIFT also could distinguish the twins. Although PSIFT could play a good performance for face recognition, it was time-consuming for pore-scale keypoints matching and it could not be applied to the actual face recognition system.

Therefore, this paper proposes a fast face recognition method based on feature matching of binary local keypoints feature. To compare with PSIFT, the method of this paper also has some experiments on the Bosphorus dataset [8]. From the experiments, we find that our method get the same face recognition accuracy, but the time to generate feature and feature matching is greatly faster than the method based on PSIFT [6].

2. A fast face recognition method based on binary local features and motion smoothness constraint

The experiment of PSIFT [4] proved that it is feasible to use local keypoint features for face recognition. However, PSIFT used a complicated method to generate the float keypoint features, and the float keypoint features matching is time-consuming. If we want to accelerate the face recognition, which is based on keypoints matching, we should find a robust matching method based on binary local features, such as Oriented FAST and Rotated BRIEF (ORB) [9]. Incorporating smoothness constraints into feature matching is known to enable ultra-robust matching. Fortunately, human face also has the motion smoothness. Motion smoothness means that if a pair of keypoints from two input image are true matched, the neighborhood of them will also have many matched pairs. If a pair of keypoints are wrong matched, the neighborhood of them will have few of matched pairs. If a pair of keypoints are wrong matched, the neighborhood of them will have few of matched pairs. Bian et. al. presented the Grid-based Motion Statistics (GMS) [10] for robust local features matching based on motion smoothness constraints. For this paper, we use the ORB feature and GMS matching method to establish a fast face recognition system based on binary keypoint features matching. The details of this paper will be introduced as follow.

2.1. ORB feature

Feature matching is at the base of many computer vision problems. ORB [9] is an efficient feature for feature matching. It is built based on FAST keypoint detector [11] and BRIEF descriptor [12]. The detector of ORB has a fast and accurate orientation component. And the descriptor of ORB is an efficient computation of oriented BRIEF features. Therefore, ORB is an efficient binary feature with good rotational and scale invariance.

2.2. Feature matching base on GMS

Then after keypoints feature being generated, feature matching is used to get the matching relationship between two images. One of most popular methods is nearest neighbor matching. The nearest neighbor matching means that it is a matched pair if the distance between a keypoint feature of image A and a keypoint feature of image B is closest. However, the matched pairs of nearest neighbor matching will contain some outliers. Many methods have been proposed to alleviate the problem with geometric information, such as RANSAC and GMS. For this paper, we use GMS to reject the false matches, so we will introduce the details of GMS. GMS is a method that converts the motion smoothness constraints into statistical measures for rejecting false matches. And it develops an efficient Grid-based score estimator that can be incorporated into a real-time feature matcher.
2.2.1. **Motion smoothness constraint.** Given a pair of images taken from different views of the same person, it is implied that a pixel in one image is identified as the same point in the other image from a feature correspondence. If the motion is smooth, neighboring pixels and features move together. As shown in Figure 1, we define the neighborhoods of some pairs of keypoints from I1 and I2 are sr1 and sr2, respectively. Motion smoothness constraint predict true match neighborhoods will have many more supporting matches than false match neighborhoods.

2.2.2. **Grid-based score estimator.** If all the match’s neighborhoods are known, motion smoothness constraint will be easily used to reject the false matches. However, it is time-consuming to calculate each feature match’s neighborhood. If the number of image features is N, the cost of scoring is O(N). Fortunately, GMS [10] proposes an efficient grid-based score estimator for motion smoothness constraint. GMS addresses that divides an image into G= 20*20 non-overlapping cells, and the neighborhood of matching pair can be seen as the cell-pair. All matches between cell-pairs deemed true are accepted, and the scores of potential cell-pairs are computed only once. This make score computation independent of feature number, and the cost is O(1). GMS groups match neighborhoods (cell-pairs). The score Sij for cell-pair {i,j} is shown as equation 1.

\[
S_{ij} = \sum_{k=1}^{K=9} |X_{ik,jk}|
\]  

(1)

Where \( |X_{ik,jk}| \) is the number of matches between cells {ik,jk} shown in Figure 2. Finally, GMS supports that it should use a threshold of Sij to divide cell-pairs into true and false sets \{T, F\} as equation 2.

\[
\text{cellpair}\{i,j\} \in \begin{cases} 
T, & \text{if } S(i,j) > \tau_t = \alpha \sqrt{n_i} \\
F, & \text{otherwise} 
\end{cases}
\]  

(2)

Where \( \alpha = 6 \) from GMS experiment, and \( n_i \) is the average (of the 9 grid-cells in Figure 2) number of features presented in a single grid-cell.
3. Experiment and analysis

To compare with PSIFT [6], we have done some experiments on Bosphorus, a public face dataset. The Bosphorus face dataset contains many images of 105 subjects with different poses and different expression. The images size of this dataset are generally 1000*1000. And we use the original image size in our experiment. Similar with PSIFT, the neutral face images of 105 subjects are defined as gallery, and the smile images, the angry images, the images of pose with 10o, 20o and 30o as the probe images. The accuracy of face recognition and the calculate time are the measurement between PSIFT and our method.

As the result of PSIFT, we use PSIFT descriptor to generate average 7238 keypoints for each input image and use nearest neighbor proportional method as feature matching algorithm. For our method, we use ORB descriptor to generate average 10000 keypoints for each image, and use the nearest neighbor method as feature matching algorithm. Then we split the 10000 keypoints as 20*20 grids, and use the GMS method to reject the false matches. After matching, we set a threshold for face recognition. If the number of matching keypoints between two input images is more than 2,000, two input images may be identified as the same person. Some results of our experiment are shown as Figure 3 and Figure 4. Figure 3 shows the matching result of the same subject with different pose and Figure 4 shows the matching result of different subjects.

Table 1 shows the accuracy and average calculate time of PSIFT and our method in Bosphorus dataset. With the similar face recognition accuracy of different expressions and different poses, our method is faster than PSIFT. Table 2 and Table 3 are show some details of matching between PSIFT and our method. PSIFT is a powerful local feature, but it only can catch average 23.8% matching rate with the same person, while the matching rate of our method is average 63.8%. Although ORB is a binary feature, it can play a good performance under GMS matching.
Figure 3. 6883 pairs matching between the same subject with different pose.

Figure 4. 68 pairs matching between the different subjects.

Table 1. The accuracy of different poses or different expression and the average calculate time.

| Method   | Smile | angry | 10 degree | 20 degree | 30 degree | Average calculate time |
|----------|-------|-------|-----------|-----------|-----------|------------------------|
| Pore-SIFT| 100%  | 100%  | 100%      | 100%      | 100%      | 1.45s                  |
| Ours     | 100%  | 100%  | 100%      | 100%      | 100%      | 0.1s                   |

Table 2. The details of the number of matching pairs of the same subjects.

| Method   | Average matching | Average number of origin | Matching rate |
|----------|------------------|--------------------------|---------------|
| Pore-SIFT| 1724.3           | 7238.4                   | 23.8%         |
| Ours     | 6388.4           | 10000                    | 63.8%         |
Table 3. The details of the number of matching pairs of different subjects.

| Method  | Average matching | Average number of origin | Matching rate |
|---------|------------------|--------------------------|---------------|
| Pore-SIFT | 51.3             | 7238.4                   | 0.7%          |
| Ours    | 46.8             | 10000                    | 0.46%         |

4. Conclusion
We propose an efficient and reliable face recognition method based on local feature matching. The method uses ORB binary feature to generate keypoint features and uses GMS method for local features matching. We have done a set of experiment in Bosphorus dataset. Compared with the face recognition method based on PSIFT, the proposed method can get the same accuracy but faster 10 times.

Acknowledgement
Foundation for Young Innovative Talents in Higher Education of Guangdong, China (No.2017KQNCX272, No.2018KQNCX389).
Foundation for Characteristic Innovation Projects in Higher Education of Guangdong, China (No.KTSCX245)

References
[1] Turk M, Pentland A. (1991) Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71-86.
[2] Belhumeur, P. N., Hespanha, J., and Kriegman, D. (1997) Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence,19, 7, 711–720.
[3] Ali, A., Hussain, S. , Haroon, F. , Hussain, S. , & Khan, M. F. . (2012). Face recognition with local binary patterns. Bahria University Journal of Information & Communication Technologies, 5(1), 46-50.
[4] Face Recognition system DLIB http://dlib.net/
[5] Swansea. Ding, C., & Tao, D.. (2015). Robust face recognition via multimodal deep face representation.  ieee transactions on multimedia, 17(11), 2049-2058.
[6] Li D, Lam K M. (2015) Design and learn distinctive features from pore-scale facial keypoints. Pattern Recognition, 48(3):732-745.
[7] Lowe D G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2):91-110.
[8] Savran A, Akarun L. (2008) Bosphorus Database for 3D Face Analysis. Biometrics and Identity Management. Springer-Verlag, Berlin.
[9] Rublee E, Rabaud V,Konolige K, et al. (2012) ORB: An efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision. Xiamen. pp. 2564-2571.
[10] Bian J, Lin W Y, Matsushita Y, et al. (2017) GMS: Grid-Based Motion Statistics for Fast, Ultra-Robust Feature Correspondence. In: IEEE Conference on Computer Vision and Pattern Recognition. Hawaii. pp. 2828-2837.
[11] Rosten E, Drummond T. (2006) Machine Learning for High-Speed Corner Detection. In: European Conference on Computer Vision. Springer, Berlin, Heidelberg. pp. 430-443.
[12] Calonder M, Lepetit V, Strecha C, et al. (2010) BRIEF: binary robust independent elementary features. In: European Conference on Computer Vision. Hersonius. pp. 778-792.