International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 10, No. 7 (2017), pp. 195-204
http//dx.doi.org/10.14257/ijsip.2017.10.7.17

Efficient Human Face Recognition Method under Subtle SIFT Features Using Optimized K-means

Zhikai Zong

School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China
zzksdu@163.com

Abstract

K-means proposed by MacQueen has been successfully and widely applied to pattern recognition and machine learning. However, the performance of k-means in human face recognition has barely been surveyed systematically. In this paper, we combine optimized k-means clustering algorithm with scale invariant feature transform (SIFT) features to improve face recognition rate. To extract SIFT features from test face image, subtle human face features will be obtained after clustering of SIFT features by optimized k-means. Human face is identified by calculating the distance between subtle features. The large scale of experiments on JAFFE and FERET face databases have been carried out to prove the effectiveness of the proposed algorithm, and compared with other methods such as k-means, SIFT features, linear discriminant analysis (LDA) and principal components analysis (PCA). The experimental results demonstrate that the high recognition rates can be obtained by the proposed method.

Keywords*: Human Face Recognition, Optimized K-means, Scale Invariant Feature Transform (SIFT)*

1. Introduction

Face recognition has received a great deal of attention over 20 years because of its wide application in the field of computer vision. However, due to that human face is not a unique and rigid object, face recognition is remaining to be a challenging problem. Numerous factors can make it easily for human face to vary under different situations. Therefore, how to select the human face feature effectively becomes crucial to enhance the accuracy of face recognition. Multiple methods have been presented for face recognition. One of the earliest attempts was proposed by Kanade [1], who extracted a vector of 16 facial parameters by the methods of image processing. In Kanade's method, the human face was represented by the sample features and recognition rate was not ideal. For the purpose of obtaining the detailed feature of human face images, Liu *et al*. [2] proposed Gabor wavelet method which computed Gabor-filter images. Meanwhile, local Gabor binary patterns (LGBP) was proposed by Choi *et al*. [3]. In their experiment, the LGBP result was decided by color histogram, and some detailed human face information was lost. Dey and Deb [4] proposed a human face recognition method based on k-means clustering extended study. It calculated the cluster values of the vector derived from each image matrix. The matching process was complexity because they failed to reduce the image's dimension. In [5], the principal components analysis (PCA) was applied to extract feature of human face and reduce the dimension with the help of covariance analysis. So as to verify the clustering effect of k-means on human face recognition, Bag *et al*. [6] proposed to extract features by the way of Eigenfaces method and identify the human face by the way of a modified k-means clustering. In [7], the supervised slow feature analysis (SSFA) was proposed to decrease dimension for face recognition. In their

ISSN: 2005-4254 IJSIP

experiment, each human face was identified by larger features, and the computational complexity was high.

In this paper, we extract the scale invariant feature transform (SIFT) features from human face. An optimized k-means method is applied to cluster the SIFT features to seek the best clustering result. Face recognition is derived from the comparison of SIFT features between test face image and image in face library. The SIFT [8] is always the focus of pattern recognition, and it can get brilliant performance in term of target recognition. K-means clustering algorithm is a typical clustering algorithm based on distance, and it has great value in data mining and machine learning. After extracting SIFT features, we use least square error function [9] to calculate the optimal group number of these characteristics, and get the optimal clustering effect. In addition, the proposed method reduces the complexity of matching between the features by the method of calculating the distance between features.

The rest of this paper is organized as follows. Section 2 briefly overviews related work of human face. SIFT method and an optimized k-means algorithm are described in Section 3. Section 4 illustrates simulation experiment and analysis of the proposed method. Conclusion and remarks on possible further work are given finally in Section 5.

2. Related Work

In this section, we will introduce the related work of human face recognition: scale invariant feature transform (SIFT) and extended study of k-means.

2.1. Scale Invariant Feature Transform

Lowe [10] proposed SIFT which extracted the invariant features from images to match the features of the same class. It mainly involves two parts [11]: feature point detector and descriptor. SIFT feature has the advantage of scale invariance. Besides, even if you change the rotation angle, image brightness or shooting angle, it can still get good detection effect.

SIFT method uses scale-space to simulate the image data of multi-scale features. Gaussian convolution kernel is the only linear nuclear to achieve the scale transformation. Therefore, the scale-space difference-of-Gaussian (DOG) is widely applied in the detection of key points from the images, which is shown as:

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)=L(x,y,k\sigma)-L(x,y,\sigma), \tag{1}$$

where $I(x,y)$ is an input image, and $L(x,y,\sigma)$ is scale space. $G(x,y,\sigma)$ is a variable-scale Gaussian, and σ is the standard deviation of normal distribution. $D(x,y,\sigma)$ is scale-space DOG. For the purpose of selecting the local maxima and minima of $D(x,y,\sigma)$ in scale-space and two-dimensional image space, each sample point is in comparison with its eight neighbors in the same scale as well as nine neighbors in the scale of the above and the below.

After determining the characteristic points in each image, a direction is calculated for each feature point. The gradient magnitude $m(x,y)$ and the orientation $\theta(x,y)$ are shown as:

$$m(x,y)=\sqrt{[L(x+1,y)-L(x-1,y)]^2+[L(x,y+1)-L(x,y-1)]^2}, \tag{2}$$

$$\theta(x,y)=\tan^{-1}\{[L(x,y+1)-L(x,y-1)]/[L(x+1,y)-L(x-1,y)]\}. \tag{3}$$

The SIFT features can be determined by the three key information which are position, orientation and scale. The face recognition is performed by matching between SIFT features.

2.2. Extended Study of K-means

K-means clustering algorithm is a typical clustering algorithm based on distance, which has good local search ability. Poomagal and Hamsapriya [12] proposed cluster analysis for similar characteristics of the data to understand these data more clearly. K-means algorithm has great value in data mining, machine learning, and pattern recognition [13]. K-means is an unsupervised clustering, and it is an indirectly clustering method based on similarity measure between samples. K-means algorithm has a host of advantages such as fast, robust, easier to understand, and relatively efficient [4]. The extended study of k-means for face recognition can be demonstrated as following steps:

Step 1. Resize the image with size of $M \times N$ into column vectors with length of $M \times N \times 1$. So, the face database is transformed into a matrix with length of $M \times N \times k$, in which k is the number of images in face database.

Step 2. Calculate the clusters value of each column vectors by k-means to form clustering matrix.

Step 3. Calculate the clusters value of test image by k-means.

Step 4. Calculate the distance of clustering values between test image and every image in face database.

This method has lots of advantages including robust, easier to understand. However, this method failed to extract human face detailed features and it didn't reduce the image's dimension. There are several aspects to improve the algorithm, such as extraction of facial features and optimized k-means.

3. Face Recognition Based on SIFT Features and Optimized K-means

In this section, efficient human face recognition algorithm will be introduced. It includes three parts: Firstly, the SIFT features are extracted from each input human face; Secondly, all the human face SIFT features are clustered with assistance of optimized k-means, and the feature matrix is obtained which contains all facial features; Thirdly, a matching computation is processed between the test image feature matrix and the feature matrix. Figure 1 shows the flow chart of human face recognition method based on SIFT features and optimized k-means.

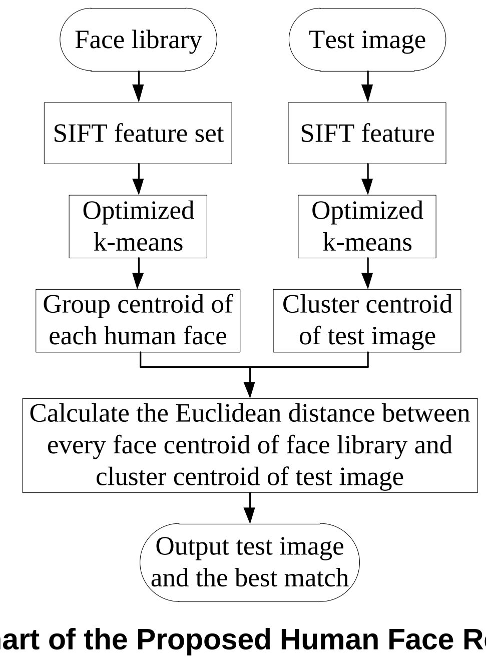


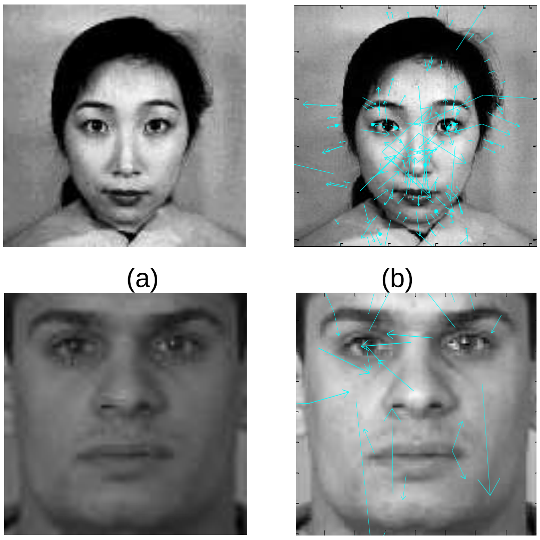
Figure 1. Flow Chart of the Proposed Human Face Recognition Method

3.1. SIFT Feature Extraction

After detecting the distinctive invariant features in every image by the SIFT method, we compute the gradient and orientation of each image sample point with the assistance of a Gaussian window. SIFT features of sample face are shown in Figure 2. From Figure 2, we can conclude that the SIFT features on each face is unique.

Table 1 and Table 2 show data of SIFT features information. From Table 1 and Table 2, the Row represents the features' abscissa, and the Column is on behalf of features' ordinate. Scale stands for the features' magnitude, and the Orientation is the name of the features' angle. Furthermore, we retain four decimal places for each feature so as to effectively differentiate each feature. In order to preserve the stability and simplicity of the clustering, we adjust the each face's SIFT features with size of 180×4 according to the specific size of every SIFT feature in JAFFE face library, and the size of each face's SIFT feature is 50×4 in FERET face library.

From the Table 1 and Table 2, we can conclude that the SIFT features of the same human face is different. A face image can be perfectly represented by these features. Therefore, it is feasible to calculate the distance between features instead of the feature matching.



(a) (b) (c) (d)

Figure 2. SIFT Features on Sample Face Images: (a) Original Image in JAFFE, (b) SIFT Features on Face Image in JAFFE, (c) Original Image in FERET, and (d) SIFT Features on Face Image in FERET

Table 1. SIFT Features on Sample Face in JAFFE

Feature	Row	Column	Scale	Orientation
1	32.2903	60.1445	6.1156	-2.2790
2	31.9567	60.0587	6.1167	-2.1954
3	30.0712	59.5736	6.1231	-1.7082
4	26.0484	58.5386	6.1366	-0.5585
5	20.9572	57.2287	6.1537	0.8715
6	16.1574	55.9938	6.1699	2.0677
7	13.0040	55.1879	6.1806	2.5163
8	11.4951	55.7316	6.2023	1.9006
⋮	⋮	⋮	⋮	⋮
180	36.5340	0.0276	0.9629	3.4371

Table 2. SIFT Features on Sample Face in FERET

Feature	Row	Column	Scale	Orientation
1	36.1233	17.3721	8.8735	-1.4656
2	34.5104	26.7621	8.3484	-1.5179
3	29.6198	51.1726	7.0527	-1.7191
4	33.3864	61.9276	5.9200	-1.2795
5	55.5052	47.5608	5.0921	0.3402
6	66.6702	32.0204	4.6842	1.6150
7	43.3501	25.0847	5.0529	1.4656
8	15.1524	24.2076	5.5607	1.0796
⋮	⋮	⋮	⋮	⋮
50	8.3933	72.4284	1.1071	-2.1478

3.2. Feature Clustering Based on Optimized K-means

When k is set to different numbers, the result of clustering is different. In other word, the value of k will decide the optimal performance of clustering. A squared error function is used to help us to define the number of k, which is shown as:

$$J = \frac{1}{k}\sum_{i=1}^{k}\left\| x^{(i)} - \mu_{k^{(i)}} \right\|^2, \tag{4}$$

where $x^{(i)}$ is the ith value of the SIFT feature, and $\mu_{k^{(i)}}$ is the ith value of SIFT features in the kth group centroid. k is the number of group centroid, and i is the ith value of every SIFT feature. K is composed of k. The initial value of k is relatively small and the value of J decreases rapidly with the increase of k. However, when k is increased to a certain value, the value of J will decline slowly. The value of k is the best choice when the graph of J is on the inflection, as shown in Figure 3.

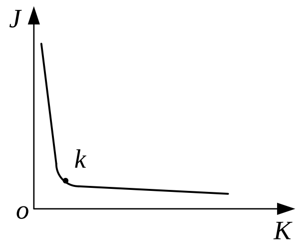


Figure 3. Relationship Curve between J and K

The SIFT features which include the number and the position of features in each image are different. In order to improve the efficiency of SIFT feature matching between face library and probe faces, the SIFT features are clustered by k-means. Each face is divided into k sub-regions, and every block has one centroid to describe the sub-region. How to choose the value of k is crucial to obtain an optimal result. Figure 4 shows the relationship curves between J and K in SIFT features. From Figure 4, it indicates that the optimal value of k is roughly same in different iterative times. Moreover, the optimal value of k is becoming more and more stable with the increase of iterative times. K-means algorithm is adopted to search the optimal results by the way of iteration. Different iterations will lead to different results. Table 3 shows the best values of k at different iterations. In order to ensure the stability of the clustering results, the initial iteration time is set to 50. Table 4 shows the gradients for the optimal value of k at different iterations. We choose the largest gradient because clustering has a better convergence at the largest gradient.

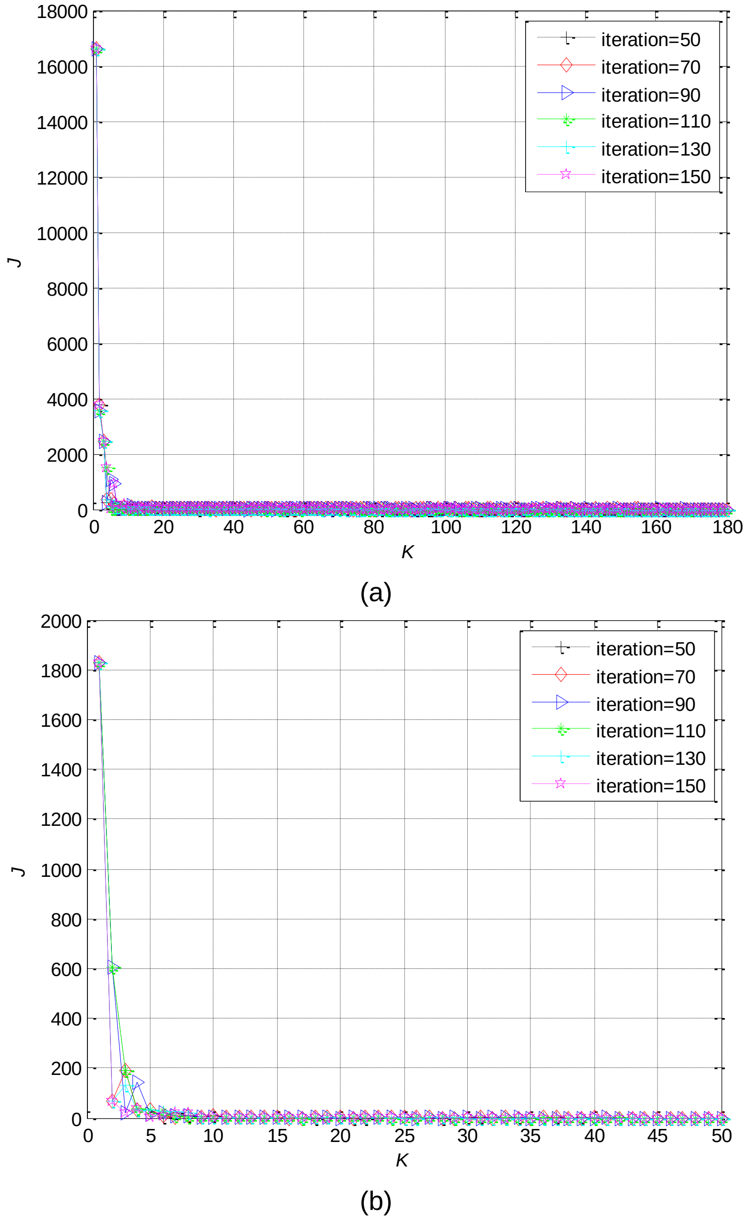


Figure 4. Relationship Curve between J and K under Different Iterations in Different Face Libraries: (a) In JAFFE and (b) In FERET

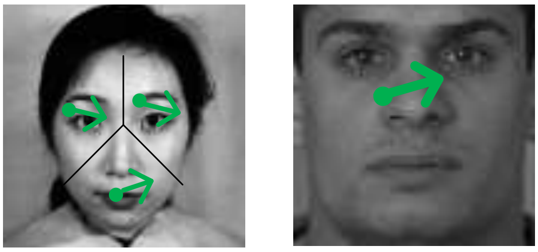
Table 3. The Best Value of k at Different Iterations in JAFFE and FERET

Iteration	50	60	70	80	90	100	110	120	130	140	150
JAFFE	4	4	3	3	3	3	2	3	3	3	3
FERET	3	1	1	1	1	2	1	1	1	1	1

Table 4. The Gradient for the Optimal Value of k at Different Iterations in JAFFE and FERET

Iteration	50	60	70	80	90	100	110	120	130	140	150
JAFFE	4	10	10	11	10	10	10	10	8	10	12
FERET	20	21	20	20	20	27	30	28	26	26	25

From Table 3 and Table 4, we can conclude that every face image in JAFFE is divided into three parts and every face in FERET is united in one part. Figure 5 shows SIFT features on human face in JAFFE and FERET after clustering. In addition, the endpoint of arrow represents the location of the SIFT feature, and the length of arrow is on behalf of the magnitude of SIFT feature. The direction of the arrow stands for the orientation of SIFT feature. In a word, every human face is represented by SIFT features.



(a) (b)

Figure 5. SIFT Features on Human Face in Different Face Libraries: (a) Sub-regions on Human Face in JAFFE and (b) Sub-regions on Human Face in FERET

3.3. Matching

After clustering, each face image is represented by k clustering centroids. Then, each face in JAFFE face library is represented by a matrix with size of 3×4 and each face in the FERET face library is represented by a matrix with size of 1×4. The process of matching is turned into calculating the distance of SIFT features between test face images and each face in face library. Euclidean distance d is adopted to calculate the distance of SIFT features, and which is shown as:

$$d_i = \sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}\left[p_i(m,n) - p_2(m,n)\right]^2}, \tag{5}$$

where $p_i(m,n)$ is the SIFT feature's centroid of ith image in face library, and $p_2(m,n)$ is the SIFT feature's centroid of test face image. M is the number of SIFT features, and N is the number of the feature's dimension. d_i is the distance between test image and the ith image in face library. When the distance between all images in the face library and test image is calculated, the image with the smallest distance is the best matching image.

4. Simulation Experiment and Analysis

The proposed algorithm is tested on MATLAB 2012a with a computer of Inter (R) Core (TM) 2 2.93GHz CPU, 4GB memory. To investigate the effect of human face recognition in our method, JAFFE and FERET are adopted as face libraries which are collected online http://www.kasrl.org/jaffe.html and http://www.frvt.org/. JAFFE database includes 213 images which contains 7 facial expressions posed by 10 female models, and the images in JAFFE are resized to 256×256. FERET database contains 14126 images of 1196 different individuals which includes illumination variation and expression variation. In addition, the images in FERET are resized to 80×80. Figure 6 shows different expression images of one individual in JAFFE database and another individual in FERET database. The algorithm is composed of the following steps:

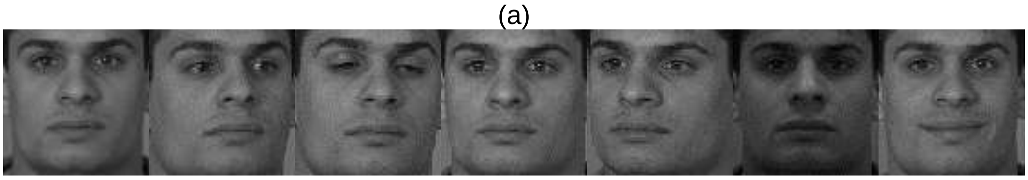
Step 1. Select face library, and extract the SIFT features of human face in face library.

Step 2. Cluster the SIFT features by the way of optimized k-means, and form the feature matrix of human face library.

Step 3. Extract the SIFT feature of test image, and compute the distance between features in test image and human face in face library.

Step 4. Identify the nearest image.

(a)



(b)

Figure 6. Images of One Individual in Different Databases: (a) In JAFFE and (b) In FERET

The JAFFE and FERET face databases are adopted to evaluate the proposed method. In the FERET face database, 100 face images are selected from the 100 different persons and in the JAFFE face database, 100 face images are selected from different persons. In FERET, we take 100 images which include 80 images from face library and 20 images of illumination variations from outside the face library. In JAFFE, we choose 100 face images from different people. From the relationship curve between J and K, we define that k is set as three in JAFFE and k is set as one in FERET. The clustering result is more stable if the iteration has the greater number. Considering the overall efficiency and accuracy, the number of iteration is set as 150 in JAFFE and the number of iteration is set as 110 in FERET.

Table 5 shows the performance of proposed method, PCA method [15], LDA method [15] and extended study of k-means method [4]. Table 6 shows the performance of proposed method, SSFA method [7], EBGM_Optimal method [14] and SIFT_GRID method [6]. In Table 5 and Table 6, the face recognition rate is between 0 and 1, and the greater value represents the better performance of face recognition.

Table 5. The Recognition Rate of Different Methods on JAFFE

Methods	Recognition Rate
PCA [15]	0.83
LDA [15]	0.82
Extended Study of K-means [4]	0.90
Proposed	**0.92**

Table 6. The Recognition Rate of Different Methods on FERET

Methods	Face Library	Illumination Variations
EBGM_Optimal [14]	0.90	0.42
SIFT_GRID [6]	0.94	0.35
SSFA [7]	0.90	0.40
Proposed	**0.96**	**0.45**

From Table 5 and Table 6, we can draw a conclusion that the proposed algorithm has better performance than other methods, and it reduces the redundancy of SIFT features matching. The high face recognition is due to the fact that detailed features are collected by SIFT and the subtle SIFT features cluster centroids are obtained by optimized k-means algorithm.

5. Conclusions

In this paper, we propose an efficient human face recognition method, which combines SIFT with an optimized k-means. Due to the fact that SIFT features of each person are different, the k-means algorithm can find the subtle features of each face and unify the number of SIFT features on each face. Feature matching is replaced by computing the distance between features, which avoids the disadvantages of a large amount of calculation in the matching process.

For further work, we will combine better feature extraction algorithm with optimized k-means algorithm to improve the rate of human face recognition under illumination variation.

References

[1] T. Kanade, "Picture Processing System by Computer Complex and Recognition of Human Faces", Kyoto University, Japan, Ph.D. Thesis, DOI: 10.14989/doctor.k1486, **(1974)**.

[2] C. J. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fished linear discriminant model for human face recognition", IEEE Transactions on Image Processing, vol. 11, no. 4, **(2002)**, pp. 467-476.

[3] E. Choi, S. W. Lee, and C. Wallraven, "Face recognition with enhance local Gabor binary pattern from human fixations", Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Seoul, South Korea, **(2012)**, pp. 863-867.

[4] T. Dey and T. Deb, "Extended study of k-means clustering technique for human face classification and recognition", Proceedings of the 1st International Conference on Electrical, Computer and Communication Technologies, Coimbatore, India, Article number: 7226023, **(2015)**.

[5] F. Mahmud, M. E. Haque, S. T. Zuhori, and B. Pal, "Human face recognition using PCA based genetic algorithm", Proceedings of the 1st International Conference on Electrical Engineering and Information & Communication Technology, Dhaka, Bangladesh, Article number: 6919046, **(2014)**.

[6] S. Bag, S. Barik, P. Sen, and G. Sanyal, “A statistical nonparametric approach of face recognition: Combination of Eigenface and modified k-means clustering”, Proceedings of the 2nd International Conference on Information Processing, Bangalore, India, **(2008)**, pp. 198-204.
[7] X. J. Gu, C. C. Liu, and Z. J. Yang, “Dimensionality reduction based on supervised slow feature analysis for face recognition”, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 7, no. 1, **(2014)**, pp. 13-26.
[8] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, “On the use of SIFT features for face authentication”, Proceedings of the 6th IEEE International Conference on Computer Vision and Pattern Recognition Workshop, New York, USA, Article number: 1640475, **(2006)**.
[9] Q. Wang, C. Wang, Z. Y. Feng, and J. F. Ye, “Review of k-means clustering algorithm”, Electronic Design Engineering, vol. 20, no. 7, **(2012)**, pp. 21-24.
[10] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, vol. 60, no. 2, **(2004)**, pp. 91-110.
[11] J. Luo, Y. Ma, E. Takikawa, S.H. Lao, M. Kawade, and B. L. Lu, “Person-specific SIFT features for face recognition”, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, USA, vol. 2, **(2007)**, pp. 593-596.
[12] S. Poomagal and T. Hamsapriya, “K-means for search results clustering using URL and tag contents”, Proceedings of the International Conference on Process Automation, Control and Computing, Coimbatore, India, Article number: 5978906, **(2011)**.
[13] M. Elshourbagy, E. Hemayed, and M. Fayek, “Enhanced bag of words using multilevel k-means for human activity recognition”, Egyptian Informatics Journal, vol. 17, no. 2, **(2016)**, pp. 227-237.
[14] L. Wiskott, J-M. Fellous, N. Krüger, and C. V. Malsburg, “Face recognition by elastic bunch graph matching”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, **(1997)**, pp. 775-779.
[15] A. M. Martinez and A. C. Kak, “PCA versus LDA”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, **(2001)**, pp. 228-233.

Author

Zhikai Zong, he was born in Shandong province, China, in 1993. He received his B.E. degree in electronic information engineering from Shandong University of Technology, China, in 2016. He is currently pursuing his M.E. degree at Shandong University, Weihai, China. His current research interests include digital image processing and computer vision.

