J. Vis. Commun. Image R. 55 (2018) 795–801

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Face-mask recognition for fraud prevention using Gaussian mixture model ☆

Ququ Chen [a,*], Lei Sang [b]

[a] School of Computer and Information, Hefei University of Technology, Hefei, China
[b] Academy of Photoelectric Technology, Hefei University of Technology, Hefei, China

ARTICLE INFO

Article history:
Received 9 May 2018
Revised 23 July 2018
Accepted 20 August 2018
Available online 22 August 2018

Keywords:
Gussian Mixture Model (GMM)
Dlib
Deep learning
Fraud prevention

ABSTRACT

With the rapid development of biometric identification technology, face recognition has been one of the most widely used as its important component. It facilitates a series of applications such as security, military, transportation, education and other fields. The demand for face feature recognition is increasing. However the current techniques still exist some deficiencies. In this paper, we proposed a face-mask recognition method for fraud prevention based on Gaussian Mixture Model. We address the problem of identifying the face and mask in the area of financial security precaution. And we show how to combine opencv with dlib to recognition face and extract it. We use Gaussian Mixture Model (GMM) to construct the model of human faces. According to this, we calculate the similarity between the face sample and the model. By analyzing and learning the features of faces, we can predict whether the image of which we test is a human face or a mask. Compared with other traditional method of face recognition, our approach has been targeted to strengthen the ability to recognize abnormal faces such as sunglasses, masks and respirator, and reduce the potential danger of these unusual faces in the security field. It is simple to be calculated and has a higher accuracy. In addition, our method have enhanced the robustness of the algorithm about mask recognition.

1. Introduction

In recent years, artificial intelligence has attracted more and more attention from the public as a great breakthrough in this field. As a new biometric technology, face recognition has become a very hot topic, the demand for this technology of people are increasing gradually. Compared with traditional identification methods, the greatest advantages of face recognition are more secure and more convenient. Human face recognition technology has unique ability of discriminate actively, which ensures that other people cannot cheat the system with some non-active photos, puppets, waxworks, etc. It can be used in security and transportation surveillance areas [2–6,11,12], because it does not required user to cooperate with biometric method [22]. The confidentiality of the operation is significantly enhanced. In addition, it adopts non-contact collection which is not invasive and easy to be accepted. Compared with other biometrics, face recognition belongs to an automatic recognition technology [2,16,20], which is fast and difficult to be perceived.

Broadly speaking, face recognition system represent a series of relative technologies, and contains four main components: collecting and detecting [30,31] images of human face, preprocessing face image, extracting face feature and recognizing face. Different face images including static images, different gestures and various expressions can be collected by camera. Extracting the video frame can also intercept the face images which people need. Various pattern features are abundant in face images, such as color features, features, template features, structural features and Haar features. To detect these, the main current method of face detection is Adaboost algorithm by picking out some of the most representative matrix face features (weakly classifier) [10,14,15,17], then training these weakly classifiers to construct strong classifiers according to the ways of weighted voting. After that a cascade classifier composed of several strong classifiers is put into a cascade structure, which can effectively improve the detection speed of the classifier. However using this method, the main problems are low classification accuracy because of the imbalance data, iteration times are difficult to set, and spending long training time. Due to various restrictions on the conditions and random interference, the vast

☆ This article is part of the Special Issue on TIUSM.
* Corresponding author.
E-mail address: ququchen8@gmail.com (Q. Chen).

https://doi.org/10.1016/j.jvcir.2018.08.016
1047-3203/

majority of cases original images cannot be directly used. Thus images are required to reprocessed including histogram equalization, light compensation, gray-level correction and so on [7,8]. On this basis, some features of faces could be extracted.

There are many methods for face recognition [13,19,33], such as geometric features algorithm. Geometrical features can be the shape of the eyes, nose, mouth, and the geometrical relation between them (such as the distance between each other). These algorithms recognize have fast speed and require small memory, but the recognition rate is low. In addition, feature-based recognition algorithms, appearance-based recognition algorithms, template-based recognition algorithms, and recognition algorithms using neural network [32,34].

Different from the existing method of face recognition, we focus on the recognition of people who are wearing masks, sunglasses and respirators. These abnormal faces can easily arouse people's vigilance in some areas of automatic recognition such as bank login system and residential security. Due to this abnormal faces have a high similarity [23,24] with normal faces, it is easy to default to them as normal faces in automatic identification system, so we would sometimes ignore the potential risks that these unusual people might cause. In this paper, we proposed a new method of face recognition. It consists of three parts: (1) Using Haar classifier in opencv to detect face in images, and then extract the face. (2) A GMM is trained for a large number of images based on the feature vectors of the human face. (3) The dlib library of deep learning algorithm is used to detect again and generate face features.

2. Related work

Clustering is often used effectively in image processing, which is used to solve the segmentation of eigenspace based on the aggregation degree of eigenspace in the extraction of image characteristics. Ng et al. [26] presented a medical image segmentation algorithm based on K-means and modified watershed segmentation algorithm. By adopting K-means clustering, this unsupervised learning algorithm was used to make a preliminary of images, but the number of the pseudo and over-segmented graphs is easily generated. Zivkovic et al. [27] put forward a typical background subsection of computer vision project. Through the analysis of commonly used method of pixel level, using the method of Gaussian mixture model to select the appropriate number of components for each pixel, and constructing a statistical model on the image of a scene without an intrusion object. On this basis, they detect an intrusion object by non-conforming parts to achieve an analysis of the image scene.

In the field of face recognition, many methods have been proposed relative to our current work. Tran et al. [29] proposed a visual observation confidence method to reduce the impact of light on face recognition. By selecting the appropriate weights in the aspects of learning characteristics such as discrimination, distance and illumination could get the optimal combination. Zhang et al. [1,21,19] proposed a new feature selection based on emotion recognition. The original acoustic features are mapped and retained by calculation of the time consumption, redundancy, etc., and then keep a suitable portion of the emotion features. This approach could improve efficiency of emotion recognition, However it is easy to lose part of information when used for face recognition. In order to represent and recognize different shapes of faces, Graham et al. [25] described a characteristic space manifold to identify features of theoretical identification based on distribution of facial feature. They proposed an improved face recognition method to enhance the process from unfamiliar to familiar face recognition. In addition, this method can be used to get general recognized face from video.

The sparse signal representation method [9,16,18] has been applied to automatic recognition of faces in recent years. Wright et al. [28] proposed a general classification algorithm for object recognition based on the sparse representation based on l1. In the case of sparse representation theory prediction, the traditional feature of feature face and Laplace face can be displayed as long as the dimension of the eigenspace is more than a certain threshold, so as to help predict how many occlusion problems can be processed by the algorithm for these unconventional features such as sampling images and random projection. It also helps to select training images to maximize the robustness against occlusion.

3. Proposed method

In this part, this paper proposed a method based on GMM and deep learning, which uses a deep learning dlib library to extract the characteristic of human face in the image to generate a 128 dimensional feature vector. As we can see in Fig. 1, we classify features of the image by GMM model, learning characteristic from training data, and using the GMM algorithm to predict the image of test. Since dlib is insensitive in identifying of people who wear masks, respirator and sunglasses, so opencv face detector is added for secondary screening. Therefore, in order to improve the robust of face recognition performance, we set up multiple parameters in opencv to adjust the detection speed on the one hand, and use dlib for face alignment and extraction features on the other hand, so as to reduce the time consumption of face detection. By using the combination of opencv detection and dlib detection, as long as one part fails to detect the face, or cannot align, it can be considered that no human face can effectively distinguish the mask or non-face image from the face, and reduce the false recognition of non-face images such as masks into images with human faces. This method is of certain significance for the face recognition in the financial security field for masks and wearing sunglasses, reducing the potential risks associated with these abnormal faces.

For the overall framework, this paper mainly includes three steps: face detection, feature extraction and model construction.

3.1. Face detection

In our method, we use opencv to detect images which can filter out some does not contain the human face or some face covered by mask and respirator. It improves the accuracy of face detection, reduce the residual rate of mask in opencv. As shown in Fig. 2, we use the haar feature for face detection. The haar feature is first proposed by Papageorigiou, it was used for face description, which is a common feature description operator in computer vision field. It can reflect the gray-scale variation of image. The difference value can be reflected by the sub-module of pixels, which can reflect the gray-scale variation of the image.

3.2. Feature extraction

After using the opencv to get the images that contains faces, we utilize dlib to exact the features of face. Dilb is a modern c++ toolkit, which included machine learning algorithms and tools, which provided a human face detection algorithm and a feature point algorithm. We use dlib face detector to import a 68-point model of the human face into the shape_predictor, to align the face to a standard pose. Then ResNet face recognition model was loaded to extract the 128-dimensional face features. ResNet is deep residual network that was first proposed by Kaiming He. By utilizing the network to learn residual, it can be more powerful than CNN in depth and accuracy. The traditional conventional layer causes problems of information defect and distortion due to the increase

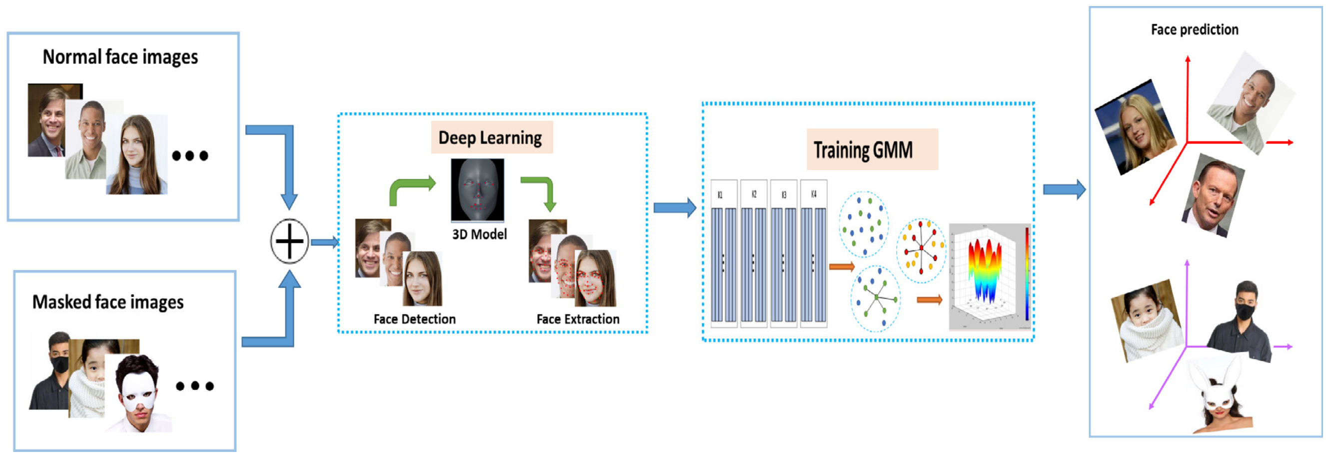


Fig. 1. Integrated framework of our face recognition approach.

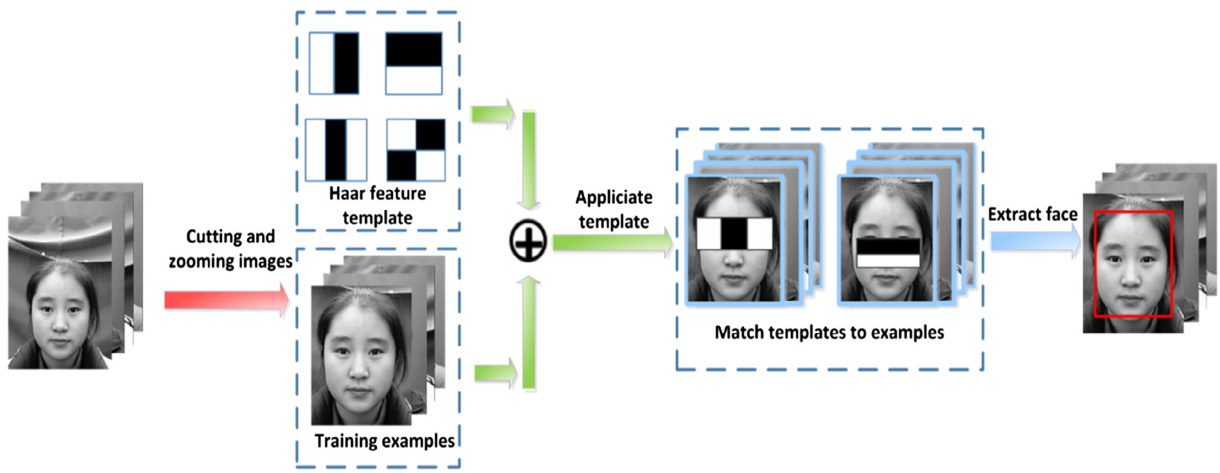


Fig. 2. Face detection process utilizing opencv detector.

of layer number during passing messages. While ResNet has many bypass brunch to protect the integrity of information by passing it around to the output.

3.3. GMM construction

Gaussian mixture model refers to the linear combination of gaussian distribution function. Theoretically, GMM can fit any type of distribution and is often used to solve the data contained in the same collection with multiple different distributions. And GMM has been applied to image segmentation, object recognition and video analysis etc. For any given set of data samples, the probability distribution of each sample data vector can be calculated according to its distribution probability, which can be classified according to the probability distribution.

When we get the multi-dimensional features of each image, store the characteristic data of N sample images, and the collection of these N sample data points $\{x_1, x_2, \ldots, x_n\}$ is subject to the gaussian distribution. GMM as a picture classifier, divides the data set into k components, corresponding to k clusters. Each component contributes p(k) to the generated data set, and a GMM can be formulated as the following objective function:

$$p(x) = \sum_{k=1}^{K} p(k)p(x|k) = \sum_{k=1}^{K} \pi_k \Pr(x|\mu_k, \Sigma_k) \tag{1}$$

where $\Pr(x|\mu_k, \Sigma_k)$ present the k component in the mixture.

For the multi-dimensional gaussian normal distribution, The density function$\Pr(x|\mu, \Sigma)$ can be represented as follows:

$$\Pr\left(x|\mu, \Sigma\right) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\} \tag{2}$$

where π_k is a weighting factor that represents the contribution of each component to the generated data set, μ is the average value of each component, and Σ is the covariance of each component. In the GMM, we use EM algorithm to solve these three parameters. First, initialize the average value, covariance and influencing factors of each cluster K.

E step:

Calculate the probability generated for each component based on the current data points.

$$\gamma(i,k) = \frac{\pi_k \Pr(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \Pr\left(x_i\middle|\mu_j, \Sigma_j\right)} \tag{3}$$

where $\gamma(i,k)$ means the posterior probability of the data point belongs to the cluster k.

M step:

According to the calculation in the above formula, the parameter value is updated constantly:

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(i,k) x_i \tag{4}$$

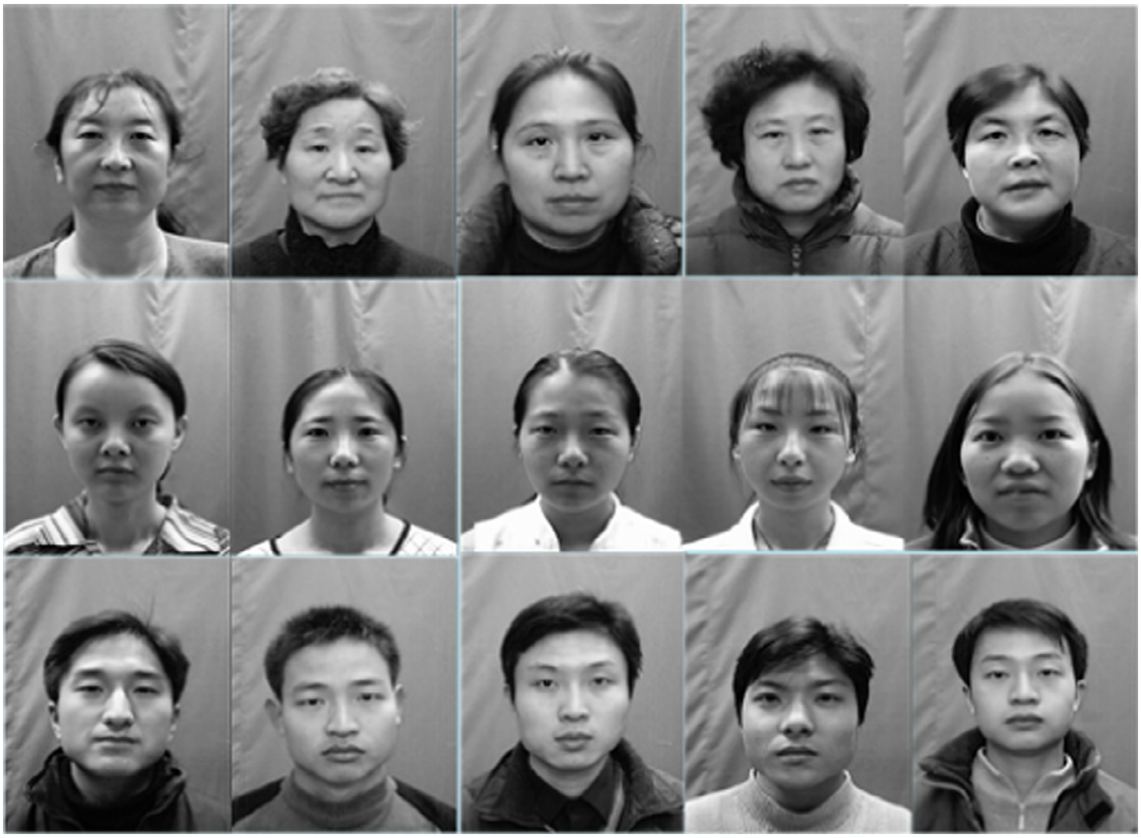


Fig. 3. some facial images in the CAS-PEAL database.

Fig. 4. some normal face images in the Real-time dataset.

$$\Sigma_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)(x_i-\mu_k)(x_i-\mu_k)^T \tag{5}$$

$$\pi_k = \frac{N_k}{N} \tag{6}$$

where $N_k = \sum_{i=1}^{N}\gamma(i,k)$ stand for the number of data points of the Kth cluster.

In view of that small probability of a single point, may small numbers can easily cause floating point numbers underflow in the computer, so we calculate the log-likelihood function:

$$\ln p(x|\pi,\mu,\Sigma) = \sum_{i=1}^{N}\ln\left\{\sum_{k=1}^{K}\pi_k \Pr(x_i|\mu_k,\Sigma_k)\right\} \tag{7}$$

Check whether the log-likelihood function is convergent (satisfying the maximum number of iterations and minimum error), and then exit the iteration, otherwise continue the E-M step.

4. Experiment

In this part, we analyze the result through several different data sets to verify the effectiveness, practicality and stability of the method. Our analysis is quantitative and qualitative. The computer hardware parameters used in our experiment are Intel (R) Xeon(R) processor and 8G memory, Windows 7 operating system, and our algorithm is implemented in Visual Studio2015.

Fig. 5. Some abnormal faces of MaskNet Dataset.

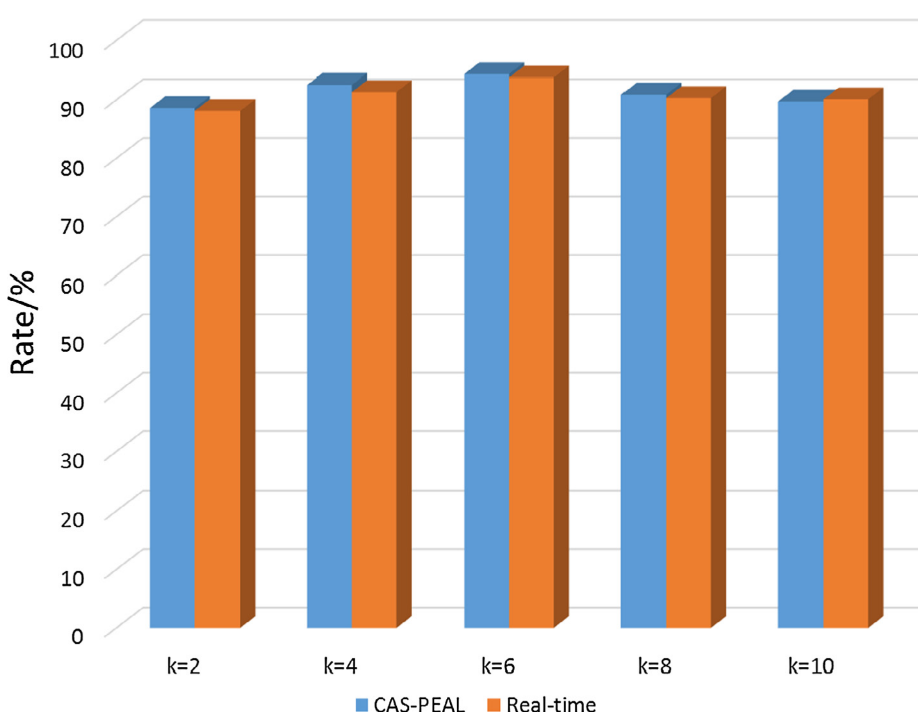


Fig. 6. The recognition performance of our approach in different when testing on the CAS-PEAL dataset and Real-time Set1.

4.1. Dataset and evaluation index

In order to verify the performance of the proposed algorithm, we use the CAS-PEAL face databases as shown in Fig. 3 and real-time face images in the financial field as shown in Fig. 4 which is constructed by ourself to conduct relevant training and testing. CAS-PEAL face databases is set by Institute Of Computing Technology Chinese Academy Of Science. It contains nearly 1040 subjects for the relevant tests, the image database is made by the Chinese academy of sciences institute of computer acquisition on China's human face image. This data set contains nearly 1040 subject in total of 30,900 images, including a variety of different gender, posture, facial expression, illumination, background and so on.

We selected 1000 random samples in the Expression and Pose data subset for the training model, containing 100 different people that each person has five different expressions and five different postures. And 300 images of each type were used for training, the rest of the images are used for testing. In the Normal data subset, 700 samples of different people with different genders and different ages were randomly selected. Among then, 500 images were used for training and 200 images were used for testing. We selected 200 images from the real-time face images in the financial area, including 1500 for training and 500 for testing. In addition, we found 400 face photos of the face being blocked (including sunglasses, masks, headscarves and respirator) as test samples.

In addition, we constructed a MaskNet Dataset as shown in Fig. 5, it contains 600 abnormal faces including people who wear sunglasses, respirator, mask and so on. These faces are regarded ae negative samples to test the accuracy of our method.

Due to the limited processing capacity of the model, to get better performance and better recognition effect of images, we did a preprocessing on these images: the original image was convert to greyscale, and the image size was adjusted to the size of 360×480.

In the performance of test, we adopt two kinds of different test images respectively, which are normal face images and blocked face images. We used three evaluation indexes that are often used in biometric signature for profiling: recognition correct rate, error acception rate and the average detection time. Recognition correct rate refers to the ratio of the number of positive samples identified as normal faces to the total number of samples.

The error acceptance rate will be the ratio of the number of normal human faces to the number of normal human faces, the smaller the better; The average recognition time, from the input to the identified test samples to the output identification results required time, the smaller the better. The error acceptance rate refers to the ratio of the number of negative samples with occluded face identify to the number of normal human faces. Average recognition time means the time required for the output identification result from the input to the test sample.

4.2. Experimental results

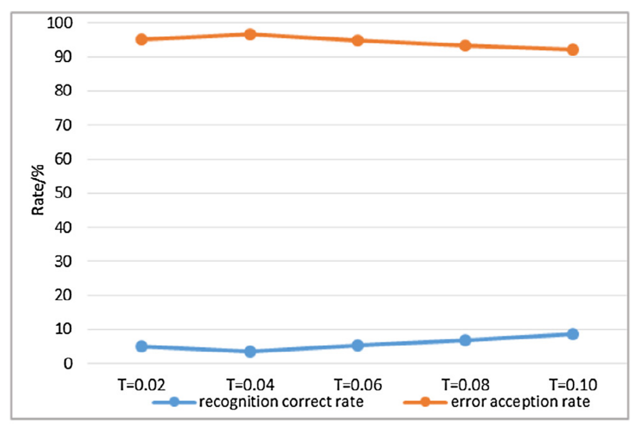
Our method uses GMM to classify image, because the training data is large and the number of multi-dimensional feature vectors is small, so the number of the model is very important. In order to improve the recognition accuracy and get a good clustering performance, we set the maximum number of iterations for 1000. Because of the Mixture Model itself is also can be arbitrarily complex, we analyzed the k = 2, 4, 6, 8, 10 of the 5 cases of algorithm performance indicators. By increasing the number of model k, we can get an obvious effect, but it also results in heavy calculation burden. The recognition performance can be seen in Fig. 6, for this two datasets, when the number of clusters k is set to 6, our approach can get a higher recognition rate.

In the stratified stage of our approach, we adopted a fixed threshold T. The threshold value is compared with the output value of the tested image in the model. When the output value is lower than the threshold, it is considered this face is abnormal and it is a mask or a face that is occluded. If the output value is above the threshold, it is considered a normal face. As shown in the Fig. 7, we analyzed the algorithm performance when T = 0.02, 0.04, 0.06, 0.08, 0.10. For this two datasets, when the threshold T is set to 0.04, our approach can get a better recognition rate.

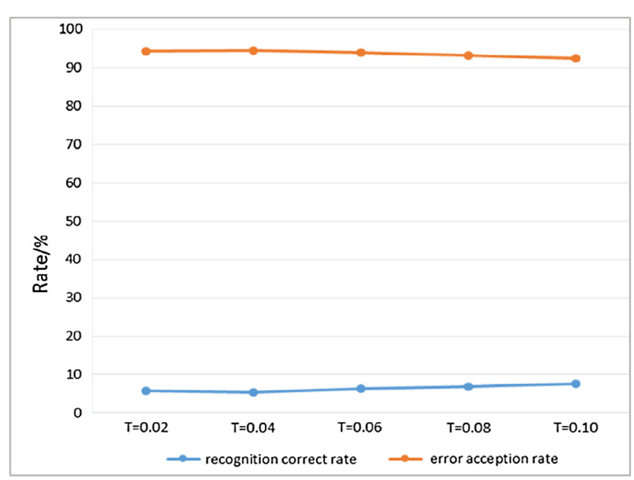
4.3. Comparison of experimental analysis

Besides above experiment, we also conduct a serious of experiments using K-means and sift method in combination with dlib or GMM to perform face recognition on the CAS-PEAL database and Real-time dataset. K-means is a typical hard clustering algorithm based on distance, regarding European distance as an evaluation index of similarity. The sift algorithm detects the local extremum points with direction information in different scales, and takes out the invariant information such as location, scale, rotation and so on, and then matches them, so as to establish the corresponding relationship between images. We verify the performance of experimental results, and we can see from the following Table 1 that out method has a superior recognition performance.

On the algorithm of time consumption, the module of image clustering, we use the GMM model classifying dlib extraction to the features of a face, because GMM in the classification task is superior to the K-means clustering method, which can learn some probability density function. It could be theoretically fitting the



(a)



(b)

Fig. 7. The recognition performance of our approach with different threshold when is tested on the CAS-PEAL database and Real-time dataset.

Table 1
Comparsion of different approaches on two databasets.

Method	CAS-PEAL databases	Real-time dataset
K-means + dlib	0.8710	0.8640
K-means + dlib + opencv	0.8850	0.8710
GMM + SIFT	0.9380	0.9340
GMM + dlib	0.9320	0.9250
GMM + dlib + opencv	0.9480	0.9390

probability distribution of arbitrary shape, and the clustering speed is faster than other algorithms. Comparing with other methods, our proposed method converges more easily.

5. Conclusion

In this paper, a real-time face detection and recognition approach based on GMM in financial anti-fraud field is proposed. We use opencv to perform a preliminary face detection in images, then we apply dlib for the second selection of images and extract of face feature which increased its accuracy. Finally we classify facial feature to adopt hierarchical framework, and the similarity between the unknown samples and training samples is calculated. This approach can improve the robustness of face recognition on occlusion change, and effectively detect the face with partially occluded face, which reduces the risk of mask fraud in the financial field. This method has certain practical significance in the future, such as residential security access control, self-service of Banks, information security and other fields.

Conflict of interest

There is no conflict of interest.

References

[1] Luming Zhang, Mingli Song, Na li, Bu Jiajun, Chun Chen, Feature selection for fast speech emotion recognition, ACM Multimedia (2009).
[2] Wei Wang, Yan Yan, Luming Zhang, Richang Hong, Nicu Sebe, Collaborative sparse coding for multi-view action recognition, IEEE Multimedia Mag. 23 (4) (2016) 80–87.
[3] Luming Zhang, Xuelong Li, Liqiang Nie, Yan Yan, Roger Zimmermann, Semantic photo retargeting under noisy image labels, ACM TOMCCAP 12 (3) (2016) 37.
[4] Luming Zhang, Xuelong Li, Liqiang Nie, Yi Yang, Yingjie Xia, Weakly supervised human fixations prediction, IEEE T-CYB 46 (1) (2016) 258–269.
[5] Luming Zhang, Meng Wang, Richang Hong, Bao-Cai Yin, Xuelong Li, Large-scale aerial image categorization using a multitask topological codebook, IEEE T-CYB 46 (2) (2016) 535–545.
[6] Luming Zhang, Yang Yang, Meng Wang, Richang Hong, Liqiang Nie, Xuelong Li, Detecting densely distributed graph patterns for fine-grained image categorization, IEEE T-IP 25 (2) (2016) 553–565.
[7] Luming Zhang, Richang Hong, Yue Gao, Rongrong Ji, Qionghai Dai, Xuelong Li, Image categorization by learning a propagated graphlet path, IEEE T-NNLS 27 (3) (2016) 674–685.
[8] Luming Zhang, Mingli Song, Qi Zhao, Xiao Liu, Bu. Jiajun, Chun Chen, Probabilistic graphlet transfer for photo cropping, IEEE T-IP 21 (5) (2013) 2887–2897.
[9] Luming Zhang, Mingli Song, Zicheng Liu, Xiao Liu, Bu. Jiajun, Chun Chen, Probabilistic graphlet cut: exploring spatial structure cue forweakly supervised image segmentation, Proc. CVPR (2013).
[10] Xiao Liu, Mingli Song, Dacheng Tao, Zicheng Liu, Luming Zhang, Bu Jiajun, Chun Chen, Semi-supervised node splitting for random forest construction, Proc. CVPR (2013).
[11] Luming Zhang, Yahong Han, Yi Yang, Mingli Song, Shuicheng Yan, Qi Tian, Discovering discriminative graphlets for aerial image categories recognition, IEEE T-IP 22 (12) (2013) 5071–5084.
[12] Luming Zhang, Yue Gao, Rongrong Ji, Qionghai Dai, Xuelong Li, Actively learning human gaze shifting paths for photo cropping, IEEE T-IP 23 (5) (2014) 2235–2245.
[13] Luming Zhang, Yue Gao, Roger Zimmermann, Qi Tian, Xuelong Li, Fusion of multi-channel local and global structural cues for photo aesthetics evaluation, IEEE T-IP 23 (3) (2014) 1419–1429.
[14] Luming Zhang, Yi Yang, Yue Gao, Changbo Wang, Yu Yi, Xuelong Li, A probabilistic associative model for segmenting weakly-supervised images, IEEE T-IP 23 (9) (2014) 4150–4159.
[15] Luming Zhang, Yue Gao, Rongrong Ji, Lv Ke, Jiale Shen, Representative discovery of structure cues for weakly-supervised image segmentation, IEEE T-MM 16 (2) (2014) 470–479.
[16] Luming Zhang, Yue Gao, Chaoqun Hong, Yinfu Feng, Jianke Zhu, Deng Cai, Feature correlation hypergraph: exploiting high-order potentials for multimodal recognition, IEEE T-CYB 44 (8) (2014) 1408–1419.
[17] Luming Zhang, Mingli Song, Yi Yang, Qi Zhao, Chen Zhao, Nicu Sebe, Weakly supervised photo cropping, IEEE T-MM 16 (1) (2014) 94–107.
[18] Luming Zhang, Mingli Song, Xiao Liu, Li Sun, Chun Chen, Bu. Jiajun, Recognizing architecture styles by hierarchical sparse coding of blocklets, Inform. Sci. 254 (2014) 141–154.
[19] Luming Zhang, Maofu Liu, Lei Chen, Lanxin Qiu, Chao Zhang, Hu Yuxing, Roger Zimmermann, Online modeling of aesthetic communities using deep perception graph analytics, IEEE T-MM (2017).
[20] Luming Zhang, Mingli Song, Xiao Liu, Bu. Jiajun, Chun Chen, Fast multi-view segment graph kernel for object classification, Signal Proces. 93 (6) (2013) 1597–1607.
[21] Luming Zhang, Yue Gao, Chao Zhang, Hanwang Zhang, Qi Tian, Roger Zimmermann, Perception-guided multimodal feature fusion for photo aesthetics assessment, ACM Multimedia (2014).
[22] Luming Zhang, Meng Wang, Liqiang Nie, Richang Hong, Yingjie Xia, Roger zimmermann: biologically inspired media quality modeling, ACM Multimedia (2015).
[23] Luming Zhang, Yi Yang, Roger Zimmermann, Fine-grained image categorization by localizing tiny object parts from unannotated images, ICMR (2015).
[24] Luming Zhang, Peiguang Jing, Su. Yuting, Chao Zhang, Ling Shao, snapvideo: personalized video generation for a sightseeing trip, IEEE T-CYB 47 (11) (2017) 3866–3878.
[25] D.B. Graham, N.M. Allinson, Characterising virtual eigensignatures for general purpose face recognition, in: Face Recognition, Springer, Berlin, Heidelberg, 1998, pp. 446–456.
[26] H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh, W.L. Nowinski, Medical image segmentation using k-means clustering and improved watershed algorithm, in: 2006 IEEE Southwest Symposium on Image Analysis and Interpretation, IEEE, 2006, pp. 61–65.
[27] Z. Zivkovic, Improved adaptive Gaussian mixture model for background subtraction, in: ICPR 2004 Proceedings of the 17th International Conference on Pattern Recognition, IEEE, 2004, vol. 2, pp. 28–31.
[28] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[29] A.T. Tran, J.Y. Kim, A. Chaudhry, H.G. Kim, Visual observation confidence based GMM face recognition robust to illumination impact in a real-world database, KSII Trans. Internet Informat. Syst. (TIIS) 10 (4) (2016) 1824–1845.
[30] X. Yao, J. Han, D. Zhang, F. Nie, Revisiting co-saliency detection: a novel approach based on two-stage multi-view spectral rotation co-clustering, IEEE Trans. Image Process. 26 (7) (2017) 3196–3209.
[31] D. Zhang, D. Meng, J. Han, Co-saliency detection via a self-paced multiple-instance learning framework, IEEE Trans. Pattern Anal. Mach. Intell. 39 (5) (2017) 865–878.
[32] J. Han, D. Zhang, G. Cheng, N. Liu, D. Xu, Advanced deep-learning techniques for salient and category-specific object detection: a survey, IEEE Signal Process. Mag. 35 (1) (2018) 84–100.
[33] J. Han, R. Quan, D. Zhang, F. Nie, Robust object co-segmentation using background prior, IEEE Trans. Image Process. 27 (4) (2018) 1639–1651.
[34] G. Cheng, C. Yang, X. Yao, L. Guo, J. Han, When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs, IEEE Trans. Geosci. Remote Sens. (2018), https://doi.org/10.1109/TGRS.2017.2783902.