Multimedia Tools and Applications (2019) 78:13297–13311
https://doi.org/10.1007/s11042-018-7002-5

Single sample face recognition via BoF using multistage KNN collaborative coding

Fan Liu[1,2] · **Sai Yang**[3] · **Yuhua Ding**[4] · **Feng Xu**[1]

Received: 31 July 2017 / Revised: 12 July 2018 / Accepted: 29 November 2018 /
Published online: 21 January 2019

Abstract
In this paper, we propose a multistage KNN collaborative coding based Bag-of-Feature (MKCC-BoF) method to address SSPP problem, which tries to weaken the semantic gap between facial features and facial identification. First, local descriptors are extracted from the single training face images and a visual dictionary is obtained offline by clustering a large set of descriptors with K-means. Then, we design a multistage KNN collaborative coding scheme to project local features into the semantic space, which is much more efficient than the most commonly used non-negative sparse coding algorithm in face recognition. To describe the spatial information as well as reduce the feature dimension, the encoded features are then pooled on spatial pyramid cells by max-pooling, which generates a histogram of visual words to represent a face image. Finally, a SVM classifier based on linear kernel is trained with the concatenated features from pooling results. Experimental results on three public face databases show that the proposed MKCC-BoF is much superior to those specially designed methods for SSPP problem. Moreover, it also has great robustness to expression, illumination, occlusion and, time variation.

Keywords Bag-of-feature · Semantic gap · Single sample per person · Sparse coding

✉ Fan Liu
faliu@hhu.edu.cn

✉ Sai Yang
yangsai166@126.com

Yuhua Ding
dingyuhua1210@163.com

Feng Xu
xufeng@hhu.edu.cn

[1] College of Computer and Information, Hohai University, Nanjing, China

[2] Nantong Ocean and Coastal Engineering Research Institute, Hohai University, Nantong, China

[3] School of Electrical Engineering, Nantong University, Nantong, China

[4] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

1 Introduction

In the past decades, face recognition has been paid more and more attention due to its great application prospect in the fields of public safety [38], transportation [36], finance [20], social media [29, 30] and so on. Discriminative feature extraction is the very first key step of face recognition, which can be generally categorized as global and local methods. Global methods treat the image as a high-dimensional vector and extract features from it. For effective and efficient global methods, subspace learning methods such as PCA [19] and FLDA [4] are adopted, which have achieved impressive results in face recognition applications. However, they are easily affected by those regions with variances in illumination, expression, occlusion etc. Comparing with the global methods, local methods have become a hot topic because of their good robustness to the local change of the image, which usually establish the feature description of the face image based on the extracted local low-level visual features, such as Gabor [22], LBP [3], SIFT [6].

In spite of the tremendous achievements, most global and local methods only work well when there are sufficient training samples for each subject. However, in many real-world applications such as identity card verification, passport verification in customs, law enforcement, surveillance or access control, only one training sample per person is available. This is so called single sample per person (SSPP) problem [32] which has become one of the greatest challenges in face recognition. Many conventional global or local methods will suffer serious performance drop or fail to work when encountering SSPP problem. This is mainly because it is difficult to distinguish the image changes caused by illumination, expression, occlusion etc. and the essential changes from different person, which leads to the semantic gap between facial features and facial identification.

Recently, the excellent performance of Bag-of-Features (BoF) methods [7, 13, 21, 27] have aroused wide interest, and been introduced into face recognition, which represents an image as a histogram of visual words. BoF extracts middle-level semantic features to weaken the semantic gap between high-level semantics and low-level features. Motivated by this point, we claim that BoF will be also suitable to solve SSPP problem. In this paper, we propose a multistage KNN collaborative coding based BoF (MKCC-BoF) method to address SSPP problem. Firstly, local descriptors are extracted from the single training face images and a visual dictionary is obtained offline by clustering a large set of descriptors with K-means. Then, we design a multistage KNN collaborative coding scheme to project local features into the semantic space, which is much more efficient than the most commonly used non-negative sparse coding algorithm in face recognition. In the k-th stage, we just use the k nearest neighbors from the visual dictionary to compute the collaborative coefficients, which further improves the computing efficiency. At the last stage, we directly use the hard vector quantization by making the coefficient of the nearest neighbor to be one and the others zero. The coding results of all stages are added together as the final coding features. To describe the spatial information as well as reduce the feature dimension, the encoded features are then pooled on spatial pyramid cells by max-pooling, which generates a histogram of visual words to represent a face image. Finally, a linear kernel based SVM classifier is trained with the concatenated features from pooling results. Experimental results on three public face databases show that the proposed MKCC-BoF not only generates well to SSPP problem but also has great robustness to expression, illumination, occlusion and , time variation.

The rest of this paper is organized as follows. We present a brief introduction to related work in the next section. Then in Section 3, we describe the proposed BoF based method in detail. Section 4 demonstrates experiments and results. Finally, we conclude in Section 5 by highlighting key points of our work.

2 Related work

How to effectively extract features from high dimensional, complex and changeable face image is the key step of face recognition. In the last two decades, subspace learning methods are the mainstream in the field of face recognition and have attracted much attentions due to their effectiveness in feature extraction and representation. Principal component analysis (PCA) [19] and fisher linear discriminant analysis (FLDA) [4] are two representative methods which respectively finds a set of optimal orthogonal basis functions to reconstruct the original signal and finds a set of optimal linear transformations to minimize the inner class divergence and maximize the divergence between classes. However, both PCA and FLDA fail to reveal the essential data structures nonlinearly embedded in high-dimensional space. To overcome this limitation, a number of manifold learning methods (e.g., ISOMAP [33], LLE [28], LPP [16], and Laplacian Eigenmap [5]) were proposed by assuming that the data lie on a low-dimensional manifold of the high-dimensional space.

Recently, the significance of feature extraction has been debated due to the excellent performance of sparse representation in face recognition. Wright et al. [39] proposed a robust face recognition via sparse representation based classification (SRC), which codes the test sample as a sparse linear combination of all training samples by $L1$ norm minimization. Then many extensions of SRC have begun to come out. Besides, it not only can be utilized for face recognition but also show great robustness in various fields such as human pose recovery [17, 43] and web image reranking [42]. To reduce the complexity of SRC, Zhang et al. [44] proposed collaborative representation-based classification (CRC) by using $L2$ norm instead of $L1$ norm. However, no matter subspace learning methods or sparse representation methods will suffer serious performance drop or even fail to work when encountering SSPP problem.

In order to address the SSPP problem, many methods have been developed during the last two decades. They can be generally classified into two categories: global methods and local methods. The global methods treat a whole image as a high-dimensional vector, which usually utilizes virtual samples or generic training set to estimate intra-personal variation. For example, Gao et al. [14] utilized SVD to decompose each face image and the obtained non-significant SVD basis images were used to estimate the within-class scatter matrix of this person approximately. Su et al. [31] proposed an adaptive generic learning(AGL) method to infer the discriminative information of the SSPP gallery set by using a generic training set. Recently, Deng et al. [12] proposed a novel generic learning method by mapping the intra-class facial difference of the generic faces to the zero vectors. They also proposed the extended sparse representation-based classifier (ESRC) [11] to make SRC feasible to SSPP problem, which applies an auxiliary intra-class variant dictionary to represent possible variation between the training and testing images. Yang et al. [41] proposed to learn the sparse variation dictionary by using the relationship between the gallery set and the external generic set.

Global methods are easily affected by those regions that are corrupted by variances in illumination, expression, and occlusion. Therefore, some local methods were proposed, which have been proven to be more robust against variations [26]. For example, Chen et al. [9] proposed the BlockFLD method by partitioning each face image into a set of blocks which treats each block as a sample from the same class and applies FLDA to the set of newly produced samples. Lu et al. [24] proposed a discriminative multi-manifold analysis (DMMA) method by learning discriminative features from image patches. The other way is to represent each patch with one feature vector. Then some famous classification techniques, such as K-nearest classifier (KNN), sparse representation based classification (SRC)

and collaborative representation based classification (CRC), can be used to predict the label of each patch, like in [26], [39] and [45]. Liu et al. [23] also proposed to use the image local structure relationship to further enhance the performance of PSRC [39] and PCRC [45].

Although local methods can lead to significant improvement in recognition rate and robustness, they still cannot distinguish the image changes caused by illumination, expression, occlusion etc. and the essential changes from different person. In other words, they cannot cross the semantic gap caused by SSPP problem. In order to fundamentally address SSPP problem, we should eliminate the semantic gap as much as possible. The direct way is to find features with semantic information. Fortunately, the excellent performance of bag-of-features (BoF) in the image classification has been introduced into face recognition in recent years, which can be regarded as a kind of middle-level semantic feature. In [21], a robust face recognition algorithm based on the block bag-of-words is proposed. Meng et al. [27] and [7] also build a bag-of-words model for face image, but they use the intensity image as local low-level features. In [40], multi-scale and multi-orientation Gabor transform are first performed on the image. Recently, Cui et al. [10] proposed a face recognition algorithm based on spatial face region description operator, which also uses intensity image to describe each image patches, and the nonnegative sparse coding method is chosen to encode each local feature. Metric learning algorithm is finally used to fuse pooling feature in each block of image. Motivated by the success of BoF, we calim that BoF with semantic information will also be suitable to solve SSPP problem.

3 The proposed approach

The overview of our face recognition using BoF with multi-stage KNN collaborative coding is shown in Fig. 1, which consists of four main steps: (1) image local feature extraction, (2) visual vocabulary construction, (3) local descriptor coding based on multistage KNN collaborative coding scheme, (4) feature pooling. The details of our algorithm are described as follows. Given a training set denoted as $R = \{(r_i, y_i)\}(i = 1, \cdots, n)$ where y_i is the class label of the i-th face image. Each image in training set is densely partitioned into a

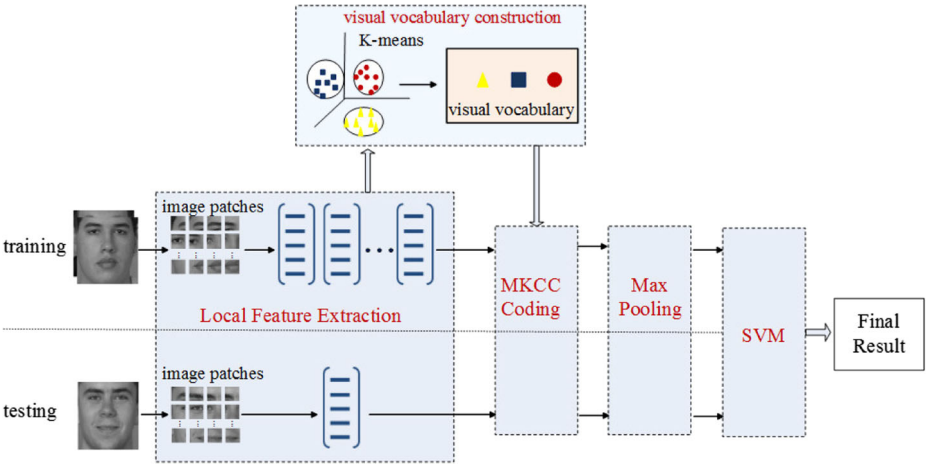


Fig. 1 Overview of our face recognition algorithm

set of patches. After the step of the local feature extraction, the set of the local features of all the training image is denoted by $X = \{x_1, x_2, \cdots, x_N\} \in R^{D\times N}$ where D is the dimension of each local feature, N is the total number of local features of training images, and the feature of each patch is extracted by SIFT descriptor. As each local feature is only a subtle description of the facial image, a large number of them are very similar. Therefore, when face images appear local changes such as illumination, facial expression and occlusion, the distance between similar local features will increase. In order to improve the robustness and discriminability of each local feature, it is necessary to use some coding algorithm to map each local feature from the low dimensional low-level visual features space to the high dimensional middle-level semantic space. It is necessary to train a complete visual dictionary offline in advance to complete the above task. To address this problem, we randomly select a subset of local features denoted as X_s from X, and use K-means clustering algorithm to cluster X_s. All the clusters form the visual vocabulary and each cluster can be regarded as a visual word which represents a specific local pattern shared by the descriptors in that cluster. The number of clusters can vary from hundreds to over tens of thousands, which determines the size of vocabulary.

Let X_r be a set of D-dimensional local descriptors extracted from each face image in image dataset, i.e. $X_r = [x_1, x_2, \cdots, x_M] \in R^{D\times M}$. Given a visual dictionary $V = [v_1, v_2, \cdots, v_K] \in R^{D\times K}$, let $c_i \in R^K$ be the coding coefficient vector of x_i. To obtain this coding coefficient vector c_i, many sparse coding methods have been proposed. However, it is very time-consuming to solve L_1 minimization. To obtain the coding efficient vector effectively and efficiently, we propose multistage KNN collaborative coding (MKCC) scheme, which utilizes L_2-norm instead of L_1-norm. And the illustration of MKCC is shown in Fig. 2. To further reduce the computing burden, we first find its k nearest neighbor visual words by euclidean metric, which is denoted as $V_k = [v_1, v_2, \cdots, v_k] \in R^{D\times k}$. And then we use V_k to code the local feature x_i by collaborative representation, which can be computed as:

$$c^* = \arg\min \|x_i - V_k c^*\|_2^2 + \lambda \|c^*\|_2 \tag{1}$$

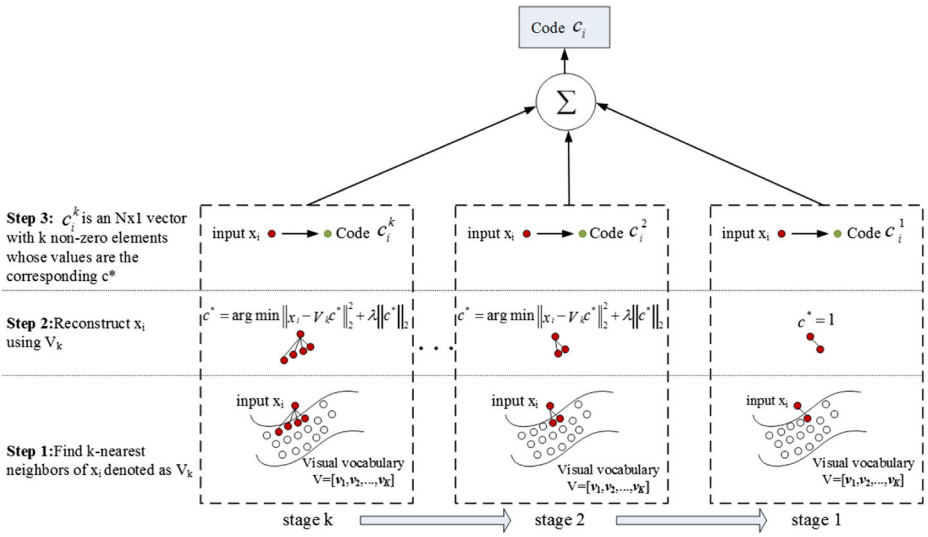


Fig. 2 The illustration of our multistage KNN collaborative coding (MKCC) algorithm

After obtaining c^*, we will get a $K \times 1$ vector c_i^k with k non-zero elements whose values are the corresponding c^*. In the next stage, the value k of KNN will become $k-1$ and we will get another $K \times 1$ vector c_i^{k-1} in the same way. This procedure is repeated until $k = 1$. It should be noticed that collaborative representation cannot work when $k = 1$. Here, we directly adopt hard vector quantization (VQ) which makes the coefficient of the nearest neighbor to be 1 and let the other elements be 0. At last, the final coding c_i of x_i is calculated by

$$c_i = \Sigma_1^k c_i^k \tag{2}$$

After the coding step is completed, the image is still represented as a set consist of $T \times M$ coded vectors. Therefore, the traditional classifier cannot be used to classify face images directly. It is necessary to compute the aggregation feature of the coded vectors to obtain a compact representation of the image content. Here, we utilize spatial pyramid method to complete pooling manipulation, which partitions an image into $2^l \times 2^l$ subregions in different scales. Let $l = 0, 1, \cdots, L$ denote the level of pyramid model, so the total levels of pyramid model is $L+1$. The illustration figure of spatial pyramid model (SPM) is shown in Fig. 3. Suppose that there are M_p encoding vectors in the pth subregion of lth level of SPM, the maximum statistical value of the coding vectors in this region is calculated as follow

$$B_{lp} = \max_{j=1,2,\cdots,M_p} c_j \tag{3}$$

The features of each sub-region from x_i in all levels are concatenated as the final representation of face image, which is denoted as B_i. Classification based on this face representation is complex due to the various facial changes like expression, illumination, occlusion etc. Support vector machine [34] is finally used to classify the images since it has high generalization performance. In case the data is linearly separable, the optimal separating hyperplane is

$$f(B_j) = sgn(\Sigma_{i=1}^n y_i \alpha_i (B_i \cdot B_j) + b^*) \tag{4}$$

where α_i is the Lagrange coefficient of each training image, B_i is the feature representation of ith training image, B_j is the feature representation of jth testing image. b^* is the threshold of classification. However, the extracted feature may be not linearly separable due to the

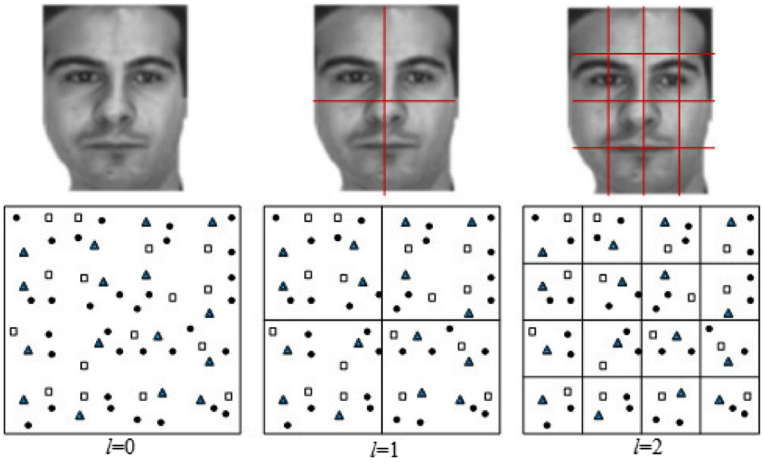


Fig. 3 Illustration of spatial pyramid model

complex facial variation. For this case, the input vectors can be nonlinearly mapped to a high dimensional feature space which is considered to be linearly separable. However, it is difficult to obtain the mapping function φ. Therefore, a kernel function $\mathcal{K}$ is utilized to compute the $\varphi^T(B_i)\cdot\varphi(B_j)$ by $\mathcal{K}(B_i, B_j)$. Then, the optimal decision surface of SVM with the kernel function is

$$f(B_j) = sgn(\Sigma_{i=1}^{n} y_i \alpha_i \mathcal{K}(B_i, B_j) + b^*) \tag{5}$$

Some popular kernel functions include linear kernel function, gaussian radial basis function(RBF), polynomial function and sigmoid function. In this paper, we use LIBSVM [8] to train a SVM classifier based on linear kernel function.

4 Experimental results

In this section, we conduct experiments on Extended Yale B [15] AR [25] and LFW [18] databases to evaluate our algorithms and compare them with several popular methods dealing with SSPP problem. These methods include AGL [31], BlockFLD [9], PCRC [45], PSRC [39], ESRC [11], SVDL [41], LGR [46] and LRA*-GL [12]. Furthermore, we also compare with some commercial projects such SeetaFace [2] and Face++ [1]. We use the gray scale of the pixels as the features for all the methods, and all the face images are resized to 80×80 in all the experiments. For patch based methods including BlockFLD, PCRC, PSRC, LGR, the patch size is fixed as 11×11 and the distance between two patch centers is 4 pixel. For our method, SIFT features are extracted by VLFeat lib [35] at single-scale from densely located patches of gray images. The patches are centered at each pixel and the fixed size is 8×8 pixels. The number of word is fixed to 1500 and the SPM is used by hierarchically partitioningeach image into 1×1, 2×2, 4×4, 8×8 and 16×16 blocks on 5 levels. Moreover, we also compare our MKCC scheme($k = 5$) with the commonly used non-negative sparse coding (NSC). All the experiments are conducted on a 2.4 GHz machine with Xeon E5-2640v4 CPU and 32G RAM. We also open 10 Matlab workers for parallel computation to improve the efficiency.

4.1 Results on Extended Yale B database

We conduct experiments with the first 30 subjects of the Extended Yale B face database, which contains 38 human subjects under 64 illumination conditions. The images of the remaining 8 subjects are used as the generic set for those generic learning methods. We use the images with the best illumination condition (0 degree azimuth and 0 degree elevation) for training and the images under other illumination conditions for testing. Some sample

Fig. 4 Sample images from the Extended Yale B database

Table 1 Recognition rate on Extended Yale B

Method	AGL	BlockFLD	PCRC	PSRC	ESRC	SVDL
accuracy	60.32	74.55	88.10	88.47	67.62	66.24
Method	LGR	LRA^*_GL	NSC-BoF	SeetaFace	Face ++	MKCC-BoF
accuracy	87.51	68.20	93.07	64.9	93.02	93.6

images from Extended Yale B database are shown in Fig. 4. Although the extreme lighting conditions make it a challenging task for most face recognition methods, the experimental results in Table 1 show that our method achieves favorable results and outperform all the other ones. It should be noticed that the recognition rate of our method is higher than the popular commercial projects SeetaFace and Face ++.

To further compare our method with SeetaFace and Face ++, we also evaluate their computing time of recognizing one image on Extended Yale B database. For SeetaFace, we use its API to extract the feature from each face image and classify the testing image by computing the cosine distance. For Face ++, we directly use it "compare API" to compute the similarity of two face images because its "search API" of the trial version is limited to search 5 face images. Then the testing image is classified into the category with highest similarity. Therefore, it needs to compute the similarity for 30 times since Extended Yale B has 30 subject to be recognized. It finally almost costs 14.85 s to recognize one image. In contrary, SeetaFace is much faster. It only consumes 0.197 s to recognize one image. However, the performance of SeetaFace is much lower than Face ++ and our method. Generally speaking, our method can achieve the best result with acceptable computing time. In addition, its computing time can also be further reduced by decreasing the size of the visual dictionary. As described above, the size of the visual dictionary is K, which also refers to the number of centers in K-means. The recognition rates and computing time under differer K is shown in Table 2. We can see that the recognition rate changes a little and even becomes a little higher when K decreases from 1500 to 50. When K is 10, the recognition rate decreases to 84.39% which is still higher than many traditional methods for SSPP problem. In addition, the computing time decreases with K decreasing.

4.2 Results on AR database

The AR face database [25] contains over 4,000 face images of 126 subjects, where 26 pictures of each subject under different facial expressions, lighting conditions and occlusions

Table 2 The impact of the visual dictionary size K

K	Accuracy(%)	Time(s)
$K = 1500$	93.6	1.85
$K = 1000$	94.97	1.53
$K = 500$	94.23	1.22
$K = 100$	94.18	1.01
$K = 50$	93.54	0.96
$K = 10$	84.39	0.93

Fig. 5 Sample images from the AR database

were taken in two sessions (separated by two weeks). In the experiments, a subset with 2500 images from 50 males and 50 females is selected, some sample images from which are shown Fig. 5.

The first 40 male and the first 40 female subjects are selected for constructing gallery and probe set and the other 20 subjects are used as the generic set of those generic learning methods. The single image of each subject with natural expression and illumination from session 1 is used for training, and the remaining images from both sessions are used for testing. Experimental results on two sessions are respectively shown in Tables 3 and 4. We can see that the proposed MKCC-BoF achieves the highest average accuracy on session 1 and the second highest on session 2. The classical non-negative sparse coding based BoF (NSC-BoF) method also achieves better results than those specially designed methods for SSPP problem. Comparing with NSC-BoF, the proposed MKCC-BoF respectively obtains 0.2% and 1.77% improvement on two sessions. Although the improvement is not very obvious, the computing efficiency of MKCC-BoF is much higher than NSC-BoF. The experimental results also show that our method is robust to expression, illumination, disguise and time variation. Although Face++ achieves the highest result on session 2, it has restriction on image size. When the image size is resized to 80×80, many face images cannot be

Table 3 Recognition rates (%) on AR database (session 1) for SSPP problem

Method	illumination	expression	disguise	illumination+disguise	avg
AGL	96.67	85.83	78.75	64.69	80.31
BlockFLD	70	82.5	75	59.69	70.52
ESRC	98.75	84.58	85.0	67.81	82.60
SVDL	98.33	86.67	85.0	69.06	83.44
PCRC	92.50	90.83	94.37	81.25	88.65
PSRC	86.25	85.83	93.13	74.69	83.44
LGR	94.17	94.17	96.88	89.69	93.1
LRA*-GL	97.92	83.75	88.75	76.25	85.62
NSC-BoF	100	99.58	99.38	98.75	99.38
SeetaFace	99.17	93.33	63.125	55.625	77.19
$Face++\{120 \times 165\}$	99.17	100	94.38	95	97.19
$Face++\{80 \times 80\}$	72.92	57.5	–	–	65.21
MKCC-BoF	100	99.58	100	99.06	99.58

Table 4 Recognition rates (%) on AR database (Session 2) for SSPP problem

Method	illumination	expression	disguise	illumination + disguise	avg
AGL	54.58	47.08	31.87	26.56	39.58
BlockFLD	72.50	53.33	61.25	40.94	55.31
ESRC	80.83	66.25	52.50	46.88	61.15
SVDL	82.92	65.83	57.50	44.69	61.67
PCRC	85.0	74.17	89.38	65.0	76.35
PSRC	77.92	67.50	81.25	55.63	68.44
LGR	85.42	81.67	93.75	79.38	83.85
LRA*-GL	85.83	66.67	70.0	60.94	70.10
NSC-BoF	96.67	93.33	95	90.63	93.54
SeetaFace	95.83	86.67	55	48.75	71.04
$Face++\{120 \times 165\}$	100	100	95.63	95	97.6
$Face++\{80 \times 80\}$	73.33	51.25	–	–	62.29
MKCC-BoF	97.92	95.42	96.25	92.81	95.31

recognized. The recognition rates under expression and illumination variation of session 1 only achieve 72.9% and 57.5%. This is because Face++ must first find face key points and extract face feature. But when the image size is too small, it cannot find face key points successfully and cannot extract face feature.

4.3 Results on LFW database

The LFW database [18] is taken under an unconstrained environment, whose images are from 5,749 individuals. In the experiments, we use the aligned version LFW-a [37] of LFW, from which 158 subjects with no less than 10 samples were gathered. Some sample images are shown in Fig. 6.

The first 80 subjects are used for evaluation, and the remaining subjects are used as the generic set. For each subject, we randomly choose one image as gallery sample and use nine images for testing. And 10 experiments are conducted to report the average recognition rates. The experimental results listed in Table 5 show that MKCC-BoF and NSC-BoF still achieve the best results. Comparing with those methods for SSPP problem, MKCC-BoF obtains nearly 10% improvement. Moreover, MKCC-BoF is still superior to NSC-BoF, which demonstrates the advantages of the proposed MKCC scheme once again.

Fig. 6 Sample images from the LFW database

Table 5 Recognition rate on LFW database

Method	AGL	BlockFLD	PCRC	PSRC	ESRC
accuracy	32.25	18.01	32.29	15.79	32.96
Method	SVDL	LGR	LRA^*_GL	NSC-BoF	MKCC-BoF
accuracy	33.18	29.36	26.90	38.99	42.68

5 Conclusion

In this paper, we try to address SSPP problem by eliminating the semantic gap between facial features and facial identification. Motivated by the success of BoF and the fact that BoF extract can extract middle-level semantic feature, we propose a multistage KNN collaborative coding based BoF (MKCC-BoF) method. Different from conventional nonnegative sparse coding based BoF methods, its computing efficiency is much faster since it has close solution. Experimental results on three public face databases show that the proposed MKCC-BoF not only generates well to SSPP problem but also has great robustness to expression, illumination, occlusion and time variation.

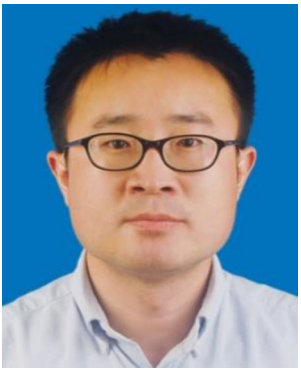
Acknowledgements This work was partially funded by National Natural Science Foundation of China under grant No. 61602150 and 61871444 , Fundamental Research Funds for the Central Universities under grant No. 2018B16214, China Postdoctoral Science Foundation funded project under grant No. 2016M600355, 2017T100323, Jiangsu Planned Projects for Postdoctoral Research Funds under grant No. 601013B, Nantong Science and Technology Project under grant No. GY12017014 and University Science Research Project of Jiangsu Province under grant No.16KJB520037.

References

1. (2012) Face++. https://www.faceplusplus.com.cn/
2. (2016) Seetafaceengine. https://github.com/seetaface/SeetaFaceEngine
3. Ahonen T, Hadid A, Pietikäinen M (2006) Face description with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Mach Intell 28(12):2037–2041
4. Belhumeur PN, Hespanha J, Kriegman DJ (1996) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. In: European conference on computer vision, pp 43–58
5. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15(6):1373–1396
6. Bicego M, Lagorio A, Grosso E, Tistarelli M (2006) On the use of sift features for face authentication. In: Conference on computer vision and pattern recognition workshop, pp 35–35
7. Cao Z, Yin Q, Tang X, Sun J (2010) Face recognition with learning-based descriptor. In: 2010 IEEE Conference on computer vision and pattern recognition (CVPR). IEEE, pp 2707–2714
8. Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):27

9. Chen S, Liu J, Zhou ZH (2004) Making flda applicable to face recognition with one sample per person. Pattern Recogn 37(7):1553–1555
10. Cui Z, Li W, Xu D, Shan S, Chen X (2013) Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3554–3561
11. Deng W, Hu J, Guo J (2012) Extended src: undersampled face recognition via intraclass variant dictionary. IEEE Trans Pattern Anal Mach Intell 34(9):1864–1870
12. Deng W, Hu J, Zhou X, Guo J (2014) Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning. Pattern Recogn 47(12):3738–3749
13. Fang Q, Sang J, Xu C (2012) Saliency aware locality-preserving coding for image classification. In: 2012 IEEE international conference on multimedia and expo (ICME). IEEE, pp 260–265
14. Gao QX, Zhang L, Zhang D (2008) Face recognition using flda with single training image per person. Appl Math Comput 205(2):726–734
15. Georghiades AS, Belhumeur PN, Kriegman DJ (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Mach Intell 23(6):643–660
16. He X, Niyogi P (2004) Locality preserving projections. In: Advances in neural information processing systems, pp 153–160
17. Hong C, Yu J, Tao D, Wang M (2015) Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval. IEEE Trans Ind Electron 62(6):3742–3751
18. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Tech. rep. Technical Report 07–49, University of Massachusetts, Amherst
19. Jolliffe IT (1986) Principal component analysis. J Mark Res 87(100):513
20. Ketcham M, Fagfae N (2016) The algorithm for financial transactions on smartphones using two-factor authentication based on passwords and face recognition. In: International symposium on natural language processing. Springer, pp 223–231
21. Li Z, Imai J, Kaneko M (2010) Robust face recognition using block-based bag of words. In: 2010 20th International conference on pattern recognition (ICPR). IEEE, pp 1285–1288
22. Liu C, Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society 11(4):467
23. Liu F, Tang J, Song Y, Zhang L, Tang Z (2015) Local structure-based sparse representation for face recognition. ACM Trans Intell Syst Technol (TIST) 7(1):2
24. Lu J, Tan YP, Wang G (2013) Discriminative multimanifold analysis for face recognition from a single training sample per person. IEEE Trans Pattern Anal Mach Intell 35(1):39–51
25. Martinez AM (1998) The ar face database. CVC technical report
26. Martínez AM (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans Pattern Anal Mach Intell 24(6):748–763
27. Meng X, Shan S, Chen X, Gao W (2006) Local visual primitives (lvp) for face modelling and recognition. In: 18th international conference on pattern recognition, 2006. ICPR, vol 2. IEEE, pp 536–539
28. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326
29. Sang J, Xu C (2012) Right buddy makes the difference: an early exploration of social relation analysis in multimedia applications. In: Proceedings of the 20th ACM international conference on multimedia. ACM, pp 19–28
30. Sang J, Xu C, Liu J (2012) User-aware image tag refinement via ternary semantic analysis. IEEE Trans Multimedia 14(3):883–895
31. Su Y, Shan S, Chen X, Gao W (2010) Adaptive generic learning for face recognition from a single sample per person. In: 2010 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 2699–2706
32. Tan X, Chen S, Zhou ZH, Zhang F (2006) Face recognition from a single image per person: a survey. Pattern Recogn 39(9):1725–1745
33. Tenenbaum JB, De Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323
34. Tzotsos A, Argialas D (2008) Support vector machine classification for object-based image analysis. In: Blaschke T, Lang S, Hay GJ (eds) Object-based image analysis. Lecture notes in geoinformation and cartography. Springer, Berlin, pp 663–677
35. Vedaldi A, Fulkerson B (2010) Vlfeat: an open and portable library of computer vision algorithms. In: Proceedings of the 18th ACM international conference on multimedia. ACM, pp 1469–1472

36. Wei X, Li H, Sun J, Chen L (2018) Unsupervised domain adaptation with regularized optimal transport for multimodal 2d+ 3d facial expression recognition. In: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). IEEE, pp 31–37
37. Wolf L, Hassner T, Taigman Y (2009) Similarity scores based on background samples. In: Asian conference on computer vision. Springer, pp 88–97
38. Woodward JD Jr, Horn C, Gatune J, Thomas A (2003) Biometrics: a look at facial recognition. Tech. rep. Rand Corp Santa Monica, CA
39. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
40. Xie S, Shan S, Chen X, Meng X, Gao W (2009) Learned local gabor patterns for face representation and recognition. Signal Process 89(12):2333–2344
41. Yang M, Van Gool L, Zhang L (2013) Sparse variation dictionary learning for face recognition with a single training sample per person. In: Proceedings of the IEEE international conference on computer vision, pp 689–696
42. Yu J, Rui Y, Tao D et al (2014) Click prediction for web image reranking using multimodal sparse coding. IEEE Trans Image Processing 23(5):2019–2032
43. Yu J, Hong C, Rui Y, Tao D (2017) Multi-task deep auto-encoder model for human pose recovery. IEEE TIE. https://doi.org/10.1109/tie.2017.2739691
44. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: Which helps face recognition? In: 2011 IEEE international conference on computer vision (ICCV). IEEE, pp 471–478
45. Zhu P, Zhang L, Hu Q, Shiu SC (2012) Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In: European conference on computer vision. Springer, pp 822–835
46. Zhu P, Yang M, Zhang L, Lee IY (2014) Local generic representation for face recognition with single sample per person. In: Asian conference on computer vision. Springer, pp 34–50



Fan Liu is currently an associate professor of HoHai University. He received his B.S. degree in network engineering from Nanjing University of Science and Technology (NUST) in June 2009. From September 2008 to December 2008, he studied in Ajou University of South Korea as an exchange student. He received his Ph.D. degree from Nanjing University of Science and Technology in January 2015. His research interests include computer vision, image processing, pattern recognition and deep learning. Dr. Liu also serves as reviewer of Information Sciences, Neurocomputing, KSII Transaction on Internet and Information Systems and Pattern Analysis and Application.

Sai Yang is currently a lecturer of Nantong University. She received his M.S. degree from School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, China, in 2010, and a Ph.D. degree from School of Computer Science and Engineering, Nanjing University of Science and Technology,China, in 2015. Her research interests include computer vision, image processing, pattern recognition and machine learning.

Yuhua Ding received his bachelor's degree in software engineering from Nanjing University of Science and Technology, Nanjing, China, in 2011, where he is currently working toward his PhD. His current research interests include pattern recognition, image processing, and subspace learning.

Feng Xu is currently a professor at Hohai University. He received his Ph.D. degree from Nanjing University in 2008. He received his B.E. and M.S. degrees from Hohai University in 1998 and 2001, respectively. His research interests include cloud computing, network information security and domain software engineering etc. He has authored over 100 journal and conference papers in these areas.