

Sparsely Encoded Local Descriptor for Face Recognition

Zhen Cui^{1,2,3}, Shiguang Shan^{1,2}, Xilin Chen^{1,2}, Lei Zhang⁴

¹ Key Lab of Intelligent Information Processing, Institute of Computer Technology, Chinese Academy of Science (CAS), Beijing, 100190, China

² Graduate University of CAS, Beijing, 100190, China

³ School of Computer Science and Technology, Huaqiao University, Xiamen, 361021, China

⁴ Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
{zcui, sgshan}@jdl.ac.cn

Abstract—In this paper, a novel Sparsely Encoded Local Descriptor (SELD) is proposed for face recognition. Compared with K-means or Random-projection tree based previous methods, sparsity constraint is introduced in our dictionary learning and sequent image encoding, which implies more stable and discriminative face representation. Sparse coding also leads to an image descriptor of summation of sparse coefficient vectors, which is quite different from existing code-words appearance frequency/(histogram)-based descriptors. Extensive experiments on both FERET and challenging LFW database show the effectiveness of the proposed SELD method. Especially on the LFW dataset, recognition accuracy comparable to the best known results is achieved.

Keywords—local descriptor; Random-projection tree; Sparse coding; face verification; face identification

I. INTRODUCTION

In the past few decades, face recognition has attracted significant attention due to its wide potential applications in public security, law enforcement, etc. Numerous methods or techniques have been presented as surveyed in [1], and considerable progress had been achieved. Currently, many face recognition systems have been able to work well under well-controlled conditions with cooperative users. However, as discovered by MBGC [2] and LFW evaluation [3], face recognition under uncontrolled environment with uncooperative users remains a great challenge. To successfully address this problem, how to represent faces plays the key role.

In the past decade, local descriptor based face representation, which models image micro-patterns, has formed a blowout [4,5,6,7,8,9,10,11,12,13,14,15], due to their robustness to extrinsic variations. A large portion of these methods are based on manually designed local patterns. For example, by combining the sign of the difference of central pixel intensity from those of its neighboring pixels, LBP [7] implicitly encodes the micro-patterns of the input image such as flat areas, spots, lines and edges. Since the sign is invariant to monotonic photometric change, LBP is robust to lighting variation to some extent. Many LBP variations or extensions have also been proposed. Zhao and Pietikäinen extended LBP to the spatial-temporal domain [12]. In order to make LBP robust to random and quantization noise in near-uniform face

regions, Local Ternary Patterns (LTP) [13] have also been proposed. By combining Gabor filtering with LBP, Local Gabor Phase Patterns (LGBP) [9] was proposed to extend LBP to multiple resolution and orientation. Later on, Histogram of Gabor Phase Patterns [8] and Local Gabor XOR Pattern [14] were further proposed to exploit the Gabor phase information. In addition, some local descriptors originally proposed for other object recognition tasks were also introduced for face recognition, such as Histogram of Oriented Gradients (HOG) [10] or SIFT [11,15]. Manually designing local pattern avoids complicated learning process. Nevertheless creating an optimal descriptor is non-trivial. One has to balance between the discriminative power and the robustness against data variance.

In contrast to the above handcrafted approaches whose patterns are predefined manually, the texton-based methods typically learn some visual primitives as code-words from a large number of local face image patches and utilize the frequency of the code-words as face representation. Considering that high-level facial semantic features consist of those low-level micro visual structures, Meng et al. proposed Local Visual Primitives (LVP) for modeling and recognition [16], which learns LVP by K-means clustering. Xie et al. [17] further applied the K-means clustering approach to patch sets sampled from the Gabor filtered face images and then quantized codes of each patch, at last concatenated block-based histograms of patterns to describe the whole face. Ahonen et al. [18] also tried K-means cluster to build local filter response codebook. More recently, Cao et al. [6] pointed that quantized codes based on K-means usually tend to have uneven distribution, so the resulting code histogram would be less informative and less compact. So they applied Random-projection tree [19] to replace K-means clustering.

Another recent progress in face recognition field is the sparse representation based method. In [20], Wright et al. proposed to recognize a face through finding its sparse coefficients with respect to the whole training set as the dictionary and seeking for the face whose samples result in the smallest reconstruction error by using their corresponding sparse coefficients. In case of multiple well-aligned samples per person, the method reports impressive accuracy, especially when faces are partially occluded. Zhang et al. [21] incorporated the face labels in the dictionary-learning stage to

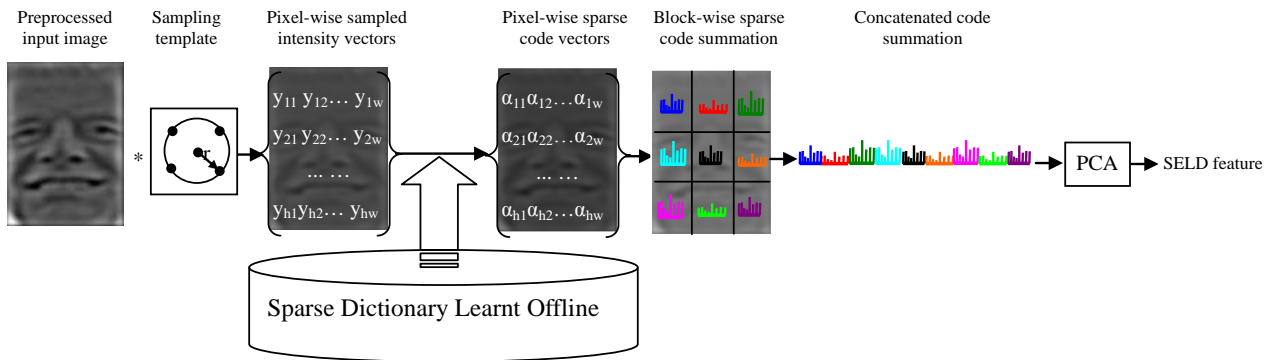


Figure 1. Overview of the proposed SELD-based face representation extraction

obtain an efficient dictionary that retains the representative power while making the dictionary discriminative. Yang et al. [22] used the image local Gabor feature for sparse representation, and proposed an associated Gabor occlusion computing algorithm to handle the occluded face image. Above methods are all holistic representation based methods, thus not as robust as local methods. Furthermore, the sparse representation method proposed in [20] can work only for scenario where each face has multiple enrolled face images. Therefore, it can not be applied to face verification scenario as evaluated by the FRGC or LFW protocol.

To address above problems, in this paper, we enhance texon-learning based local descriptor method by introducing sparse coding, thus propose Sparsely Encoded Local Descriptor (SELD). Simply speaking, in our SELD method, sparsity constraint is introduced during the local visual primitive dictionary learning and sequent image encoding, which is distinctly unlike K-means clustering or Random-projection tree as in previous methods [6, 16, 17]. As is validated recently by many researchers, sparsity implies more discriminative power and stableness of the representation.

Another big difference of our SELD method from previous texon-based methods [6, 16, 17] is that, our description is sparse coefficient vector based, rather than code-words frequency (or histogram) based. Specifically, during the image encoding stage, the coefficient vector of the sparse coding is computed at each image position, and then summed together to form the local descriptor of some image blocks. Compared with frequency (or histogram) based methods, coefficient vector is similar to some soft clustering, thus implies more robustness to variations in image appearance.

Compared with sparse representation method in [20], our SELD method is a more general face representation. As a face descriptor, our method can be easily applied to face verification and face identification with single sample per person, which are impossible for methods like in [20].

The proposed SELD method is extensively validated by experiments on two face databases: the Labeled Faces in the Wild (LFW) [3] which is designed for unconstrained face verification, and the FERET database [23] which is used for face identification. Especially on the LFW, besides the comparisons of our method with previous methods based on K-means or Random-projection tree, we also compare with

existing state-of-the-art approaches that reported best known results on LFW. Comparable accuracy is achieved by our methods.

II. SPARSELY ENCODED LOCAL DESCRIPTOR

In this section, we first present the flowchart of the proposed SELD method for face recognition. Then we describe the critical steps of our method in detail, including how to learn the sparse dictionary and how to sparsely encode a face image.

A. Overview of SELD for face recognition

As mentioned above, our method is essentially an enhanced texon-based method. Therefore, it follows similar idea to bag-of-words method. The main difference lies in the sparsity constraint in dictionary learning and the non-frequency based descriptor. Intuitively, the overall flowchart of the proposed SELD-based face representation method is illustrated in Fig.1 and explained as follows.

As shown in Figure 1, we first align and normalize the original images geometrically and filter it using a DoG filter to remove both high-frequency noise and low-frequency illumination variations. Then, at each pixel, an intensity vector is formed by sampling its neighbor pixel's intensity according to a pre-defined sampling template. In the next step, the intensity vector at each pixel is sparsely encoded with the offline-learned sparse dictionary under non-negative constraint, which generates a sparse code vector, i.e., the sparse coefficient vector. With these sparse code vectors computed, the face image is spatially partitioned into some blocks, and the sparse code vectors in each block are summed together to form a descriptor of the block. Next, the accumulated vectors of all the blocks are concatenated together to form a single vector, which is finally reduced in dimensionality by principal component analysis (PCA) to generate the SELD feature of the input face image. For face recognition or verification, cosine similarity of two SELD features can be used to match two face images.

B. Sparse Dictionary Learning

Research on general over-complete dictionaries mostly commenced over the past decade and is still intensely ongoing. Such dictionaries bring the prosperity in the definition of a signal representation. Given an over-complete dictionary matrix $D=[d_1, d_2, \dots, d_k] \in \mathbb{R}^{n \times K}$, $K > n$ that contains K prototype

signal-atoms, a signal $y \in R^n$ can be represented as a sparse linear combination of these atoms.

In this paper, we use the K-SVD algorithm [24] to train the over-complete dictionary. K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the data. Formally, given a training set with N samples, K-SVD's objective function is

$$\min_{D, X} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \forall i, \|x_i\|_0 \leq T_0 \quad (1)$$

where $X=(x_1, x_2, \dots, x_N)$ with $x_i \in R^n$ are the sparse coefficient vectors for training sample y_i , k is the number of code-words in the dictionary, and $\|\cdot\|_0$ is the l_0 norm.

The K-SVD algorithm has two stages: in the first stage, D is fixed, and the above optimization problem is a search for sparse representation with coefficients summarized in the matrix. It may be solved by any pursuit algorithm; the second stage is updating the dictionary together with the non-zero coefficients. In this stage, the algorithm updates each column in the dictionary, d_k , and the coefficients, x_R^i , the i -th row in X . The objective function (1) can be re-written as

$$\|Y - DX\|_F^2 = \left\| \left(Y - \sum_{i \neq k} d_i x_R^i \right) - d_k x_R^k \right\|_F^2 = \|E_k - d_k x_R^k\|_F^2 \quad (2)$$

Then SVD is applied to E_k and get $E_k = U \Delta V$. Then, we choose the first column of U and the first column of V multiplied by Δ_{11} as the updated d_k , x_R^k respectively.

With the above K-SVD method, our sparse dictionary is learned by the following steps:

- 1) Preprocess each normalized image in the training face image set by DoG filtering.
- 2) Sample patches of size $p \times p$ pixels from DoG filtered images to form the patch set S . If we have N training images and sample c patches from each image, there are $c \times N$ patches in S . All sampled patches are normalized to zero mean and unit length.
- 3) For patch set S , K-SVD algorithm is utilized to construct the sparse dictionary $D_{n \times K}$ where K is the number of codewords in the dictionary.

In the above learning algorithm, one of the problems might be how many patches should be sampled for training. In principle, it seems that we should sample as densely as possible to obtain a large number of patches. However, this implies very time-consuming K-SVD. Fortunately, we empirically find that only thousands of patches are sufficient for our purpose. So, the patches can be sampled rather sparsely in each image.

C. SELD-based face representation

After the sparse dictionary D is learned, we then describe in this section how to utilize it to extract SELD feature for any input testing face image. As shown in Fig.1, given an input image already normalized and filtered by DoG, we first sample patches centered at each pixel and normalize the sampled intensity vectors to zeros mean and unit length. Then, we apply

sparse coding to encode the sampled intensity vectors to sparse code vectors. Formally, let the sampled intensity vector at pixel (i, j) be y_{ij} . Its sparse code vector α_{ij} is then computed by the following optimization:

$$\alpha_{ij} = \min_{\alpha} \|y_{ij} - D\alpha\|_F^2 + \lambda \|\alpha\|_1 \quad \text{s.t.} \quad \alpha(l) \geq 0 \quad (3)$$

where $\|\cdot\|_1$ is the l_1 norm. As shown in (3), in our implement, non-negative constraint is introduced to guarantee all the entries of the sparse coefficient α non-negative. The reason we impose this constraint is that intuitively we need an additive combination of the code-words to represent each patch. This is also consistent with our sequent summation of the sparse code vectors in each image block.

After encoding, the input image is converted to ‘‘sparse code’’ map. The encoded image is then divided into an $m \times n$ grid of blocks. Then, we add all the nonnegative codes in each block to form one sparse code vector for this block. Next, the accumulated vectors of all the blocks are concatenated together to form a single vector describing the whole face image.

If we use the concatenated vector as the final descriptor, the dimension of the resulting face feature may be very high. A high-dimensional feature not only results in curse of dimensionality but also large complexity in memory and computation. Therefore, Principle Component Analysis (PCA) is further applied to further reduce the dimensionality and obtain the final SELD feature.

With the extracted SELD feature, many metrics can be used to compute the similarity or distance between two face images for face verification by threshold or identification by nearest neighbor. In this paper, we select the most commonly used cosine similarity.

III. EXPERIMENTS

In order to evaluate the proposed approach, we carry out extensive experiments on LFW benchmark [3], where we not only compare our method with previous methods based on K-means or Random-projection tree but also compare with existing state-of-the-art approaches that reported best known results on LFW. Finally, we also compare our method with K-means method for face identification on FERET database [23].

A. Experimental setting

The LFW benchmark is designed for unconstrained face verification with face images containing large variations in pose, age, expression, race and illumination. There are two evaluation modes proposed by the LFW organizer: the image-restricted and the image-unrestricted training mode. This paper only considers the restricted mode. Under this mode, the whole standard testing set consists of ten subsets and each subset contains 300 same-person pairs and 300 different-person pairs. The performance of an algorithm is measured by a 10-fold cross validation procedure. The final average recognition rate serves as the evaluation criterion. For face identification, we used FERET database and its evaluation protocol to evaluate our method.

In all experiments, DoG filters with $\sigma_1=1.0$ and $\sigma_2=2.0$ are used. The size of the sampling template is set to 5×5 . Default

dictionary size is set to 256. The PCA preserves 98% of the total energy.

B. Face verification evaluation on LFW

The original size of each image in LFW is 250×250 pixels. All face images are cropped to 80×150 pixels just by simply cutting out the center of the aligned version images provided by Wolf et al. [25]. In the block-wise SELD feature extraction stage, the face images are divided into 5×10 blocks to obtain 50 summed sparse code vectors.

Experiment 1: comparison with K-Means and Random-projection tree

The first experiment aims to validate the discriminative power and stableness of the proposed SELD by evaluating on LFW according to the LFW image-restricted evaluation mode. We compare the proposed sparse dictionary learning method with previous methods, i.e., K-means and Random-projection tree [19]. In the experiment, 500 images are randomly selected from the LFW training set to train the dictionary. Note that we find the size of training set has little influence on the final performance.

The mean accuracy curves of the three methods are shown in Fig.2, with the horizontal axis different number of code-words in the dictionary. Meanwhile, the ROC curves of the three methods are also plotted in Fig.3 when the number of code-words is 256. From these two figures, it is clear that the proposed method outperforms the other previous methods in terms of both mean accuracy and ROC. Especially, from the ROC curves, we can see that, our method works impressively better than the other two methods especially when false positive rates are small. Please note that, we have tried our best to rigidly implement the Random-projection tree algorithm as in the [19], but its performances are still slightly inferior to the K-means, which seems *slightly* different from results in [6].

In addition, from Fig.2, we can find that the performances of the three methods all increase with the increase of the code number. However, to balance the computational cost, we select 256 as the default code number in our following experiments.

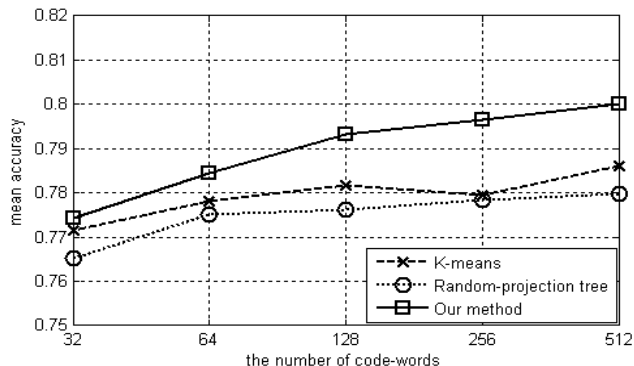


Figure 2. Performance comparison vs. learning method. We studied the mean accuracy of the learned descriptors using three learning method: K-means, Random-projection tree and the proposed sparse coding with different code number. Note that here we do not use PCA in order to reflect the comparison essentially.

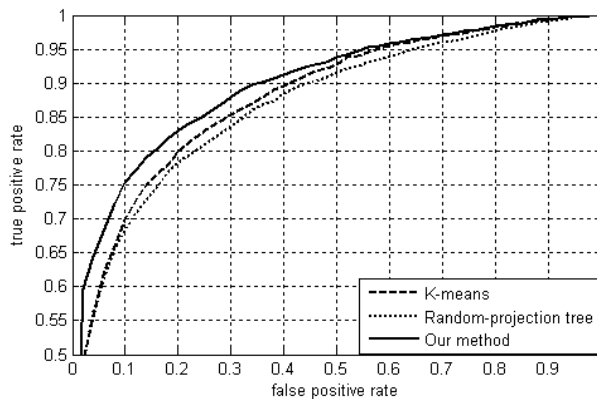


Figure 3. Demonstrate the effects of three different encoding methods in terms of ROC curves.

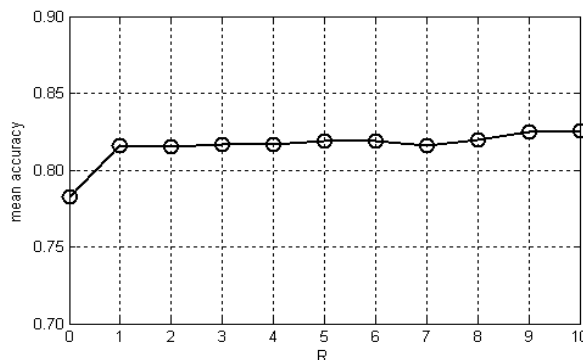


Figure 4. The effects of removing the first R dimensions of SELD feature generated by PCA

Experiment 2: effect of removing some leading PCA features

In the second experiment, we investigate the effects of removing the first R dimension from the SELD feature obtained by PCA. The reason that we focus on this point lies in the fact that, there are large variations in the LFW images, which implies the leading eigenvectors encode mostly variations in lighting, pose, and other large variances rather than those in identity. Therefore, we guess removing some of the leading eigenvectors should lead to performance improvement. The results are plotted in Fig.4. From the figure, it is clear that our above-mentioned guess is validated: by removing the first several dimensions, about 4 percents’ improvement can be achieved.

Experiment 3: fusion of multiple block-wise SELD

In above experiments, a fixed block partitioning method, i.e., 5×10 blocks, is used to extract single SELD feature. Intuitively, we have other choices to partition the image and extract multiple block-wise SELD features. Thus, when matching a pair of images, multiple similarities can be computed and fused together by sum rule or SVM. Following this idea, we tried five different block partitioning modes: 5×10, 4×8, 3×6, 2×4, 1×10, which is similar to the hierarchical spatial pyramid structure [26]. We name this method as “Multiple block-wise SELD fusion” in short.

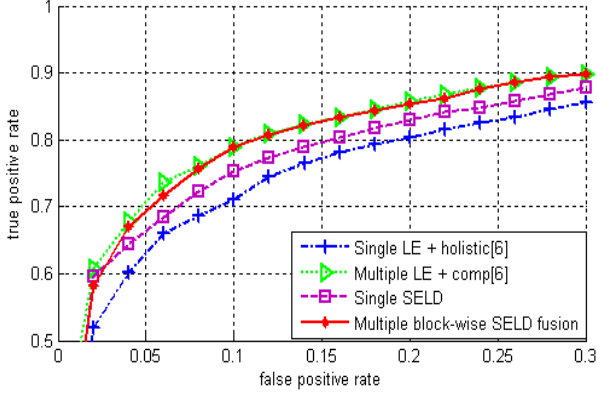


Figure 5. Our face verification methods compared with one of the state-of-the-art methods in [6], which learn dictionary based on Random-projection tree.

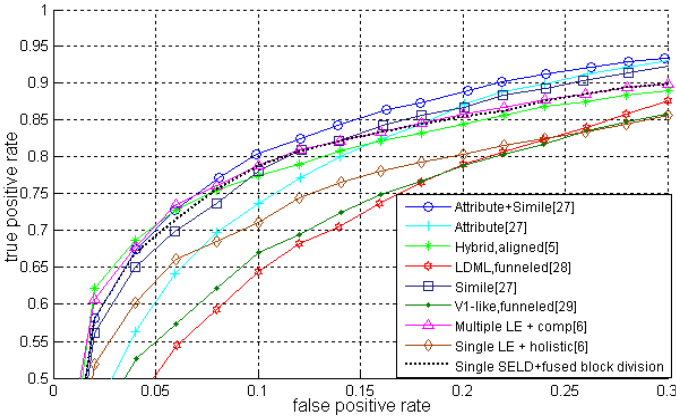


Figure 6. Face verification comparison on the LFW benchmark in restrict protocol.

In Fig.5, above method and single SELD are evaluated on LFW under image-restricted mode and compared with one of the best known method on this testing [6]. In Fig. 5, the method “Single LE+holistic” means using the single best LE to represent the holistic face, where the encoder is Random-projection tree. The method “Multi LE+comp”, reporting the best performance in [6], not only divides the holistic face into 9 components by component-level face alignment, but also fuses four different encoded local descriptors to fed into linear SVM. It also exploits pose-adaptive adjustment. In contrast, our “Multiple block-wise SELD fusion” method only exploits single sampling template and single dictionary. And the sparse coding is also conducted only once. So, it is evidently more elegant than “Multi LE+comp”.

From Fig.5, it is clear that “Multiple block-wise SELD fusion” performs impressively better than “Single SELD”, which implies that different block partitioning strategies are complementary mutually. From Fig.5, the proposed “Single SELD” consistently outperforms “Single LE” by a 4~5 percents improvement. Our “Multiple block-wise SELD fusion” is also comparable to “Multiple LE+comp”. However, as mentioned above, “Multiple LE+comp” in [6] is more complicated than our “Multiple block-wise SELD fusion” method.

Comparison with other best known results on LFW

To better validate our method, we also compare our methods with other previous state-of-the-art approaches [5,6,27,28, 29] on the same LFW evaluation, as shown in Fig. 6. From the figure, it is clear that our method is among the best ones. In addition, it is worth pointing out that, besides the training data in LFW, the method in [27] utilized an external large-scale datasets for feature extraction and classifier designing.

C. Face identification on FERET

LFW evaluation focuses on face verification. In this section, we perform experiments on the FERET dataset [23] to verify the performance of our approach for face identification.

According to the FERET evaluation protocol, algorithms were evaluated against different categories of images including some variations such as lighting change, people wearing glasses, and the time between the acquisition dates of the database image. This database consists of one standard gallery (1196 images of 1196 subjects) and four probe sets: Fb (1195 images of 1195 subjects), Fc (194 images of 194 subjects), Duplicate I (722 images of 243 subjects) (abbreviated as DupI), and Duplicate II (234 images of 75 subjects) (abbreviated as DupII).

On this database, we align and normalize the face image into 80×88 pixels. 300 frontal images are randomly selected from the FERET training CD as the training set to learn the dictionary. In our method, the sampling patch size is 5×5 pixels, and the face images are divided into 6×6 blocks.

The results of our method are compared with K-means and LVP method [16] in Table 1. In order to compare with LVP, we apply the histogram intersection as similarity measure. From Table 1, we can get several major observations. First, all method have almost equal performance on Fb dataset. Second, the SELD method is superior to local encoded method based K-means, 8, 4 and 4 percentage points improvement on Fc, DupI and DupII dataset respectively. Third, our method also significantly outperforms LVP method [16] except on Fb dataset.

D. Discussion

As evaluated in above experiments, the proposed SELD method works impressively better than similar method based on K-means or Random-projection tree. So, what is the source of the performance gain? To answer this question, we need to analyze the main different between our method and previous ones. As mentioned above, the main difference lies in two points: 1) sparsity constraint is introduced in our method during the dictionary learning, as well as during the image encoding using the dictionary; 2) block-wise summation of the sparse coding coefficients are used rather than the appearance frequency (histogram) of the code-words.

TABLE I. RECOGNITION PERFORMANCE ON THE FERET DATABASE

Descriptor	Fb	Fc	DupI	DupII
LVP [16]	0.97	0.70	0.66	0.50
K-means	0.95	0.76	0.64	0.58
SELD	0.95	0.84	0.68	0.62

The key of the first difference is “sparsity”, which has been proved in recent literature very effective for both representation and discrimination. Compared with K-means or Random-projection tree which learn local visual patterns that most frequently appears, sparse dictionary learning focuses more on the effective representation of the local patches, which endow the code-words more representation power.

The second difference actually reflects the difference of exclusive model selection and additive model summation. Give a sampled patch, traditional methods always encode it to one single code-word by vector quantization. However, our method encodes it to a vector associated with all the code-words. In other words, traditional methods use single code-words to express a patch, while our method expresses it with a linear combination of all the code-words, which we believe can lead to a more stable representation.

IV. CONCLUSION

In this paper, we propose sparsely encoded local descriptor (SELD) for robust face recognition. Unlike traditional K-means or Random-projection tree based dictionary learning method, sparse dictionary is leaned by introducing sparsity constraint. The encoding procedure is also changed from exclusive code-word assignment to sparse coding. Correspondingly, code-words frequency based descriptor is replaced by block-wise summation of sparse coding coefficients vector. The above characteristics endow the proposed SELD better performance, as validated by the LFW evaluation. Results comparable to the best known results are achieved.

In this paper, the proposed SELD method is only validated on face recognition problem. Nevertheless, SELD is not limited to face recognition, since it is not specially designed for face recognition. Therefore, we will apply it to other possible applications, e.g., object categorization.

ACKNOWLEDGMENT

This paper is partially supported by Natural Science Foundation of China under contracts No.61025010, and No.60833013; National Basic Research Program of China (973 Program) under contract 2009CB320902; and Hi-Tech Research and Development Program of China under contract No.2009AA01Z317.

REFERENCES

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Comput. Surv.*, vol. 35, no. 4, pp.399–458, 2003.
- [2] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O’ Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada1, J. A. Scallan III, and S. Weimer, “Overview of the Multiple Biometrics Grand Challenge”, in *Proc. International Conference on Biometrics*, 2009.
- [3] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” *University of Massachusetts, Amherst, Technical Report 07-49*, 2007.
- [4] G. Hua and A. Akbarzadeh, “A robust elastic and partial matching metric for face recognition,” in *Proc. ICCV*, 2009.
- [5] Y. Taigman, L. Wolf, T. Hassner, and I. Tel-Aviv, “Multiple One-Shots for utilizing class label information,” in *Proc. BMVC*, 2009.

- [6] Z. Cao, Q.Yin, J.Sun, and X.Tang, “Face recognition with learning-based descriptor,” in *Proc. CVPR*, 2010.
- [7] T.Ahonen, A.Hadid and M.Pietikäinen, “Face description with local binary patterns: application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (12) (2006) 2037–2041.
- [8] B. Zhang, S. Shan, X. Chen, W. Gao, “Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition,” *IEEE Transactions on Image Processing* 16 (1) (2007) 57–68.
- [9] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, “Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition,” in *Proc. ICCV*, 2005.
- [10] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, “Face recognition using HOG–EBGM,” *Pattern Recognition Letters*, Vol.29, pp.1537-1543, 2008.
- [11] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [12] G. Zhao and M. Pietikäinen, “Local binary pattern descriptors for dynamic texture recognition,” in *Proc. of ICPR*, volume 2, pages 211–214, 2006.
- [13] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under ~~diff~~ lighting conditions ,” in *International Workshop on Analysis and Modeling of Faces and Gestures*, 2007.
- [14] S. Xie, S. Shan, X. Chen, J. Chen, “Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition,” *IEEE Transactions on Image Processing*, vol.19, no.5, pp1349-1361,May.2010.
- [15] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, “On the Use of SIFT Features for Face Authentication,” in *Computer Vision and Pattern Recognition Workshop (CVPRW’06)*, 35, 2006.
- [16] X. Meng, S. Shan, X. Chen, W. Gao, “Local visual primitives (LVP) for face modeling and recognition,” in *International Conference on Pattern Recognition (2006)* 536–539.
- [17] S. Xie, S. Shan, X. Chen, X. Meng, and W. Gao, “Learned local gabor patterns for face representation and recognition,” *Signal Processing*, 89(12):2333–2344, 2009.
- [18] T. Ahonen and M. Pietik’ainen, “Image description using joint distribution of filter bank responses,” *Pattern Recognition Letters*. 30(4):368–376, 2009.
- [19] Y. Freund, S. Dasgupta, M. Kabra, and N. Verma, “Learning the structure of manifolds using random projections,” in *Proc. NIPS*, 2007.
- [20] J. Wright, A.Yang, A. Ganesh, S.Sastry, and Y.Ma, “Robust face recognition via sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 210-227, 2009.
- [21] Q. Zhang and B. Li, “Discriminative K-SVD for dictionary learning in face recognition,” in *Proc. CVPR*, 2010.
- [22] M. Yang and Lei Zhang, “Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary,” in *Proc. ECCV* 2010.
- [23] P. Phillips, H. Moon, S. Rizvi, P. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (10) (2000) 1090–1104.
- [24] M. Aharon, M.Elad, and A.Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on signal processing*, 54(11):4311,2006.
- [25] L.Wolf, T.Hassner, and Y.Taigman, “Similarity scores based on background samples,” in *Proc. ACCV*, 2009.
- [26] J. Yang, K. Yu, T. Huang, “Supervised translation-invariant Sparse Coding,” in *Proc. CVPR*, 2010.
- [27] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar, “Attribute and Simile classifiers for face verification,” in *Proc. ICCV*, 2009.
- [28] M.Guillaumin, J. Verbeek, C. Schmid, I. LEAR, and L.Kuntzmann, “Is that you? Metric learning approaches for face identification,” in *Proc. ICCV*, 2009.
- [29] N.Pinto, J.DiCarlo, and D.Cox, “How far can you get with a modern face recognition test set using only simple features,” in *Proc.CVPR*,2009.