Two-stage Patch-based Sparse Multi-value Descriptor for Face Recognition

Riqiang Gao [1], Wenming Yang *[2], Xiaoling Hu [3], Qingmin Liao [4]

Shenzhen Key Lab. of Information Sci&Tech, Shenzhen Engineering Lab. of IS&DRM
Department of Electronic Engineering, Graduate School at Shenzhen
Tsinghua University, China
* corresponding author: yangelwm@163.com

***Abstract*—In this paper, we propose Two-stage Patch-based Sparse Multi-value Descriptor (TPSMD), a generalization of Sparse Linear Regression Binary method. The TPSMD makes two contributions. First, the multi-value strategy introduces user-specified parameters to improve the binarization, which makes our method more discriminant and less sensitive to noise. The multi-value strategy is a comprise between the simplification and discrimination. Second, the two-stage patch-based strategy contains two independent patch-segmentations for the face image. In the first stage, according to the Multi-value strategy we obtain the discriminative local descriptor based on small patches. In the second stage, we calculate weights for larger patches, and the discriminative face regions, such as eyes and month, are strengthened by the weights. The Two-stage strategy considers local similarity in the first stage and global differences in the second one. Extensive experiments on Extended Yale B and FERET show that our method outperforms state-of-the-art methods.**

***Index Terms*—Face recognition, SLRB, Multi-value Descriptor, Two-stage Patch-based, TPSMD**

I. Introduction

Face recognition is one of the most popular and challenging problems in computer vision. Many representative methods, such as SVM [1] and SRC [2], have received good results in controlled condition. However, face recognition in uncontrolled environment is still challenging.

Some researchers normalize illumination through the traditional image processing methods, such as Gamma Intensity Correction [3]. Georghiades et al. [4] learn the model of face images under varying illumination, but the model needs a lot of training samples. Some Local feature descriptors, such as Local Binary Pattern [5], Local Ternary Pattern [6] and Sparse Linear Regression Binary[7], are widely used in illumination-robust face recognition.

Patch-based methods are widely used in face recognition. Zhu et al. [8] propose a Multi-scale Patch method, which indicates that patches of different scales have complementary information for classification. In [9], Gao et al. propose the Regularized Patch-based Representation to solve the Single Sample Per Person problem. The method of [10] operates the face alignment and recognition at the same time. And Ding et al. [11] solve the Pose-Invariant with patch-based method.

LBP is a local descriptor of texture which is widely used in robust face recognition [5]. LBP could tolerant the monotonic illumination variations and its complexity of calculation is low. However, LBP is sensitive to noise especially when neighbor pixels are similar with the center pixel. And Tan et al. [6] propose the LTP to improve LBP. Based on two assumptions: Locally Linear Consistency Assumption [12] and Lambertian Reflectance Model, Yang et al. [7] propose the Sparse Linear Regression Binary (SLRB). SLRB is superior to LBP and LTP in illumination-robust face recognition. However, binarization makes SLRB sensitive to noise and less discriminant against non-illumination images. And the SLRB [12] only considers the similarities of intra-patch, without considering the differences between different patches. However, the effects of different parts are not the same, and Ahonen et al. [13] prove that different face regions make different contributions for face recognition.

In this paper, we propose a novel Two-stage Patch-based Sparse Multi-value method. The Multi-value strategy removes the influence of illumination, while retaining more useful information. It makes the recognition discriminative. In the Two-stage Patch-based strategy, we segment the face images into patches twice independently. The first segmentation is based on Locally Linear Consistency Assumption, which considers the similarities of local patch. And because of that different regions make different contributions to recognition, the second segmentation allocates patches with different weights.

The rest of this paper is organized as follows. Section II briefly recalls the SLRB, and proposes the Multi-value descriptor. Section III would describe our model in detail. We conduct extensive experiments to test our model in Section IV and summarize the paper in Section V.

II. Sparse Multi-value Descriptor

In [7], Yang et al. propose the Sparse Linear Regression Binary to remove illumination. For the given the face image, each $N \times N$ patch has $(N-2)^2$ center pixels $f^{(k)}, k = 1, \ldots, (N-2)^2$. The center pixels are formed as a column vector:$f = \left[f^{(1)}, f^{(2)}, \ldots, f^{((N-2)^2)}\right]^T$. For each patch, we

978-1-5090-5316-2/16/$31.00 © 2016 IEEE

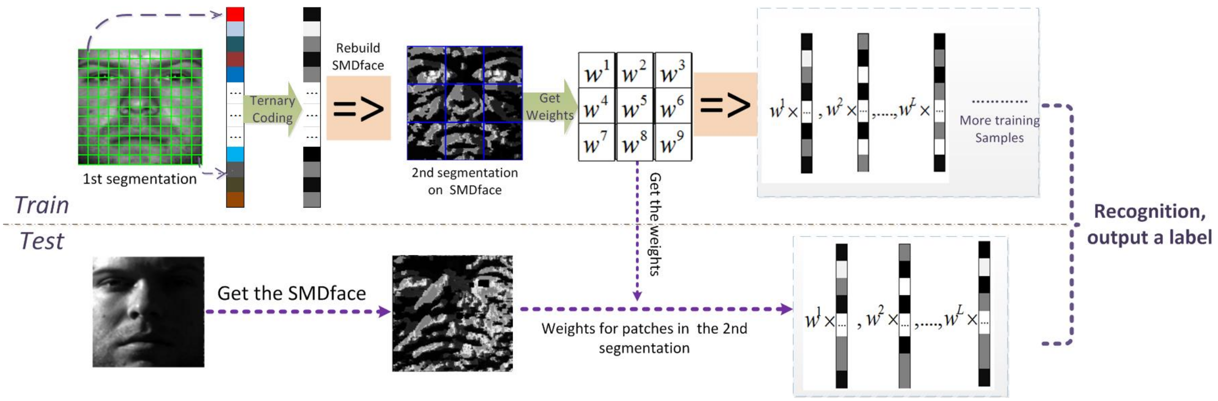

1st segmentation
Ternary Coding
Rebuild SMDface
=>
2nd segmentation on SMDface
Get Weights
w^1 w^2 w^3 w^4 w^5 w^6 w^7 w^8 w^9
=>
$w^1\times$, $w^2\times$,....., $w^L\times$
More training Samples
Train
Test
Get the weights
Recognition, output a label
Get the SMDface
Weights for patches in the 2nd segmentation
$w^1\times$, $w^2\times$,....., $w^L\times$

Fig. 1. Overview of our model. In the training phase, we get the Sparse Multi-value Descriptor in the first patch-segmentation and get the weights in the second one. The TPSMD is obtained by multiplying the Sparse Multi-value Descriptor and the weight, and descriptors in different parts may get different weights. In the test phase, we treat the test image as same as training samples, and give the same weights of training samples.

have:

$$f^{(k)} = \sum_{j=1}^{8} \alpha_j f_j^{(k)} + \varepsilon_k = x_k^T \alpha + \varepsilon_k \tag{1}$$

where $f_j^{(k)}$is the intensity of the j-th surrounding pixel of the k-th center pixel, $x_k = \left[f_1^{(k)}, f_2^{(k)}, \ \ldots, f_8^{(k)}\right]^T$ is surrounding pixels vector and $\alpha = [\alpha_1, \alpha_2, ..., \alpha_8]^T$ is coefficient vector. For $k = 1, \ldots, (N-2)^2$ of Eq. (1), we could obtain:

$$f = X\alpha + \varepsilon \tag{2}$$

where $f = \left[f^{(1)}, f^{(2)}, ..., f^{(N-2)^2}\right]^T$ and $X = \left[x_1^T, x_2^T, ..., x_{(N-2)^2}^T\right]^T$. We use the Sparse Representation [2] to optimize the coefficient vector α of Eq. (2). And the obtained vector β in Eq. (3) is the so-called SLRB:

$$\beta_j = \begin{cases} 1 & \alpha_j > 0 \\ 0 & otherwise \end{cases} \tag{3}$$

SLRB has proven to be highly discriminative for classification and it is resistant to illumination. SLRB uses sparse linear regression, which reduces the types of feature. And the binarization makes simplification further. However, the simplification causes information-loss. In [7], the experiments show that SLRB would be weaker than LBP or LTP when the illumination of images is not strong.

In this section, we introduce the Multi-value strategy in detail. Actually, The Multi-value strategy is a generalization of binarization, and it is a compromise between simplification and discriminative. However, experiments show that time cost of Multi-value is almost the same as that of SLRB.

Based on Eq. (2), the Multi-value method is described as follows:

$$\beta_j = \sum_{i=1}^{N} c_i \cdot sign\left(\alpha_j - c_i\right) \tag{4}$$

where

$$sign\left(x\right) = \begin{cases} 0 & x \le 0 \\ 1 & x > 0 \end{cases}$$

Sparse Multi-value Descriptor (SMD) is less sensitive than SLRB, and keeps the ability of de-illumination. Taking the ternary as an example, we can obtain:

$$\beta_j = c_1 \cdot sign\left(\alpha_j - c_1\right) + c_2 \cdot sign\left(\alpha_j - c_2\right) \tag{5}$$

where c_i is user-specified, which makes the the ternary codes more resistant to noise. Now, we determine the two values mainly according to experience.

III. Two-stage Patch-based Sparse Multi-value Descriptor

In this section, we introduce Two-stage Patch-based strategy in the first, and then combine the Multi-value and Two-stage Patch-based strategies as TPSMD.

The way of segmenting images is illustrated in Fig. 2. The green lines indicate the first segmentation, where we get the local descriptor based on Locally Linear Consistency Assumption. The blue lines indicate the second one, where we assign different weights to each patch according to the Fisher Separation Criterion (FSC) [14]. In these two stages, the patch-segmentations are independent.

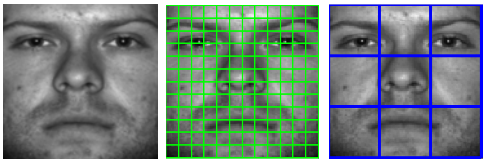
We evaluate the discrimination of different patches based on FSC. Given C different samples, we compute the mean m_{in}^l and variance σ_{in}^l for intra-class of l-th patch:

$$m_{in}^l = \frac{1}{C}\sum_{i=1}^{C}\frac{2}{N_i\left(N_i-1\right)}\sum_{k=2}^{N_i}\sum_{j=1}^{k-1} D\left(F_{i,j}^l, F_{i,k}^l\right) \tag{6}$$

$$\left(\sigma_{in}^l\right)^2 = \frac{2}{N_{in}}\sum_{i=1}^{C}\sum_{k=2}^{N_i}\sum_{j=1}^{k-1}\left(D\left(F_{i,j}^l, F_{i,k}^l\right) - m_{in}^l\right)^2 \tag{7}$$

where $F_{i,j}^l$ represents the feature of the l-th patch of j-th sample in class i. N_i represents the number of samples in class i, and $N_{in} = \sum\limits_{i=1}^{C} N_i\left(N_i-1\right) - 2$.

Similarly, we calculate the mean m_{ex}^l and variance σ_{ex}^l for inter-class of the l-th patch.



(a) Original Face (b) Stage 1 (c) Stage 2

Fig. 2. The two stage of Patch-segmentation. Generally the patch size of stage 1 is much smaller than that of stage 2. And the two stages are independent.

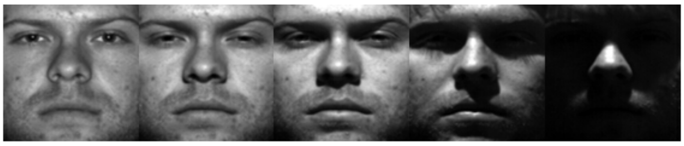


Fig. 3. Images in 5 subsets of the same person. These images are well aligned, and the illumination conditions get worse for subset 1 to subset 5.

TABLE I
THE ACCURACY IN SUBSETS OF EYB(SSPP)

Method	S1	S2	S3	S4	S5
LBP	1*	1*	0.969*	0.610*	0.349*
LTP	1*	1*	0.978*	0.766*	0.584*
SLRBl	1*	1*	0.870*	0.851	0.811
SMDl	1	1	≥**0.901**	≥**0.875**	≥**0.832**
TPSMDl	1	1	≥**0.916**	≥**0.894**	≥**0.843**
SLRBo	1	1	0.870	0.808	0.657
SMDo	1	1	≥**0.910**	≥**0.845**	≥**0.712**
TPSMDo	1	1	≥**0.917**	≥**0.888**	≥**0.826**

According to the Fisher Separation Criterion, we obtain the weight of l-th patch:

$$w^l = \frac{\left(m_{in}^l - m_{ex}^l\right)^2}{\left(\sigma_{in}^l\right)^2 + \left(\sigma_{ex}^l\right)^2} \tag{8}$$

Afterwards, we regularize the w^l with the following rule:

$$w^l := \frac{w^l - \min\{w\}}{B}, B = \frac{sum\{w\}}{n} - \min\{w\} \tag{9}$$

where n is the number of patches. To make the face recognition more robust, we make the w^l approximately to be an integer:

$$w^l = f\left(w^l - t_1\right) + f\left(w^l - t_2\right) \tag{10}$$

where

$$f(x) = \frac{1}{1 + e^{-\lambda \cdot x}}$$

.

When λ is large, the function f is similar to the $sign$. The sigmoid function f is used to restrain oversized weights, and it can partially suppress the noise. With the operation of Eq. (10), the weights can be calculated by generic database, and generic database is important to SSPP problem [15].

We assign bigger weights to those blocks that are discriminative, and weaken or abandon the others. This strategy makes full use of the discriminative patches and gets rid of the noise. The distance of gallery and probe samples could be written as the weighted sum of different patches:

$$D\left(F_g, F_p\right) = \sum_{l=1}^{M} w^l D\left(F_g^l, F_p^l\right) \tag{11}$$

As Fig. 1 shows, the TPSMD combines the Multi-value and Two-stage Patch-based strategies for face recognition. From the first segmentation we get the discriminative feature descriptor based on the Multi-value strategy, and we obtain robust weights for different locations of the face in the second one. The size of patches in these two stages are different, and the weights are used for the patches of the second stage.

IV. EXPERIMENT

In this section, we verify the effectiveness of our method on different databases: Extended Yale B and FERET. Our experiments are based on SSPP problem, and we get the weights from generic database. We use the optimization toolbox SPAMS [16] to solve the sparse estimation problems. There are several algorithms in SPAMS, and we use the $mexOMP$and $mexLasso$ to make a contrast. In our experiments, SMD represents Sparse Multi-value Descriptor and TPSMD represents Two-stage Patch-based Sparse Multi-value Descriptor.

The Extended Yale B database [17] was collected by Yale university, and it contains 38 people. In the experiments, we mainly use the face images of each sample under 64 different illuminative conditions. As Fig. 3 shows, we divide all the images into 5 subsets according to the degree of illumination. We use the Hamming Distance for classification.

The recognition results on Extended Yale B are shown in Table I, Table II and Fig. 4. The Two-stage Patch-segmentation and Multi-value strategies have improved the accuracy, especially under the condition of $mexOMP$. The results in Table II show that SMD and TPSMD are competitive when compared with state-of-the-art methods in Single Sample Per Person problem.

We compare the running time among SLRB, SMD and TPSMD. The experiments are carried out using MATLAB2013a on a 3.30 GHz computer with 8GB RAM. The results are showed in Table III. The running time of these 3 methods are almost the same. When we use the function of $mexOMP$, the running time would decrease.

TABLE II
THE ACCURACY IN SUBSETS OF EYB(SSPP)

Method	ESRC [15]	PNN[18]	PCRC[8]	SVDL[19]
Accuracy	0.679	0.675	0.778	0.850
Method	LGR[20]	SLRB	SMD	TPSMD
Accuracy	0.866	0.906	**0.922**	**0.932**

TABLE III
THE RUNNING TIME OF EYB(MS)

Method	SLRB	SMD	TPSMD
mexLasso	36.1	36.6	36.5
mexOMP	30.4	30.3	30.6

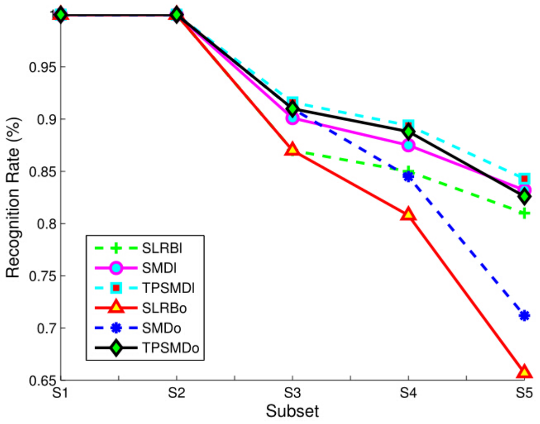


Fig. 4. Face recognition on Extended Yale B. Our methods has improve the results of SLRB, especially when using the $mexOMP$.

TABLE IV
THE ACCURACY IN SUBSETS OF FERET(SSPP)

Subset	fb	fc	Dup I	Dup II	Average
LBP	0.764	0.598	0.519	0.543	0.606
SLRB	0.859	0.830	0.734	0.671	0.774
SMD	**≥0.881**	**≥0.856**	**0.754**	**≥0.680**	**≥0.793**
TPSMD	**≥0.901**	**≥0.878**	**0.770**	**≥0.704**	**≥0.813**

FERET database contains 14051 images with multi-pose, illumination and different age of 1196 individuals. The gallery set is composed of positive-face, neutral-illumination and normal-expression images. The probe set contains 4 subsets: subset Fb with expression, subset Fc with illumination, subset Dup I with small time interval and subset Dup II with large time interval.

The results in FERET are shown in Table IV. The SLRB is proved to be competitive under illumination. And SMD and TPSMD are more discriminative than SLRB under expression, illumination and age conditions.

V. CONCLUSION

In our work, we have proposed a new algorithm for robust face recognition. Our method TPSMD is a generalization of SLRB, where we add two robust strategies: Multi-value and Two-stage patch-segmentation. The Multi-value strategy could maintain the ability of de-illumination, and makes the recognition more discriminative. The Two-stage patch-segmentation strategy takes the local and global features into consideration. According to the theories and experimental results, our method is proved to be competitive in face recognition.

VI. ACKNOWLEDGEMENT

This work was supported in part by the Natural Science Foundation of China under Grant No.61471216 and in part by the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under Grant No.JCYJ20150831192224146, No.JCYJ20150601165744635 and ZDSYS20140509172959974

REFERENCES

[1] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 688–694.

[2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[3] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*. IEEE, 2003, pp. 157–164.

[4] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 2, pp. 218–233, 2003.

[5] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[6] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[7] Z. Yang, Y. Wu, W. Zhao, Y. Zhou, Z. Lu, W. Li, and Q. Liao, "A novel illumination-robust local descriptor based on sparse linear regression," *Digital Signal Processing*, vol. 48, pp. 269–275, 2016.

[8] P. Zhu, L. Zhang, Q. Hu, and S. C. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 822–835.

[9] S. Gao, K. Jia, L. Zhuang, and Y. Ma, "Neither global nor local: Regularized patch-based representation for single sample per person face recognition," *International Journal of Computer Vision*, vol. 111, no. 3, pp. 365–383, 2015.

[10] L. Zhuang, T.-H. Chan, A. Y. Yang, S. S. Sastry, and Y. Ma, "Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment," *International Journal of Computer Vision*, vol. 114, no. 2-3, pp. 272–287, 2015.

[11] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face recognition," *Image Processing, IEEE Transactions on*, vol. 24, no. 3, pp. 980–993, 2015.

[12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[14] R. Parry and I. Essa, "Feature weighting for segmentation," *Proc. ICMIR*, pp. 116–119, 2004.

[15] W. Deng, J. Hu, and J. Guo, "Extended src: undersampled face recognition via intraclass variant dictionary," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1864–1870, 2012.

[16] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Foundations and Trends® in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.

[17] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 643–660, 2001.

[18] R. Kumar, A. Banerjee, B. C. Vemuri, and H. Pfister, "Maximizing all margins: Pushing face recognition with kernel plurality," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2375–2382.

[19] M. Yang, L. Gool, and L. Zhang, "Sparse variation dictionary learning for face recognition with a single training sample per person," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 689–696.

[20] P. Zhu, M. Yang, L. Zhang, and I.-Y. Lee, "Local generic representation for face recognition with single sample per person," in *Computer Vision–ACCV 2014*. Springer, 2014, pp. 34–50.