LETTER

# Weighted Voting of Discriminative Regions for Face Recognition*



**Wenming YANG**[†], ***Member***, **Riqiang GAO**[†a)], ***and*** **Qingmin LIAO**[†], ***Nonmembers***

**SUMMARY** This paper presents a strategy, Weighted Voting of Discriminative Regions (WVDR), to improve the face recognition performance, especially in Small Sample Size (SSS) and occlusion situations. In WVDR, we extract the discriminative regions according to facial key points and abandon the rest parts. Considering different regions of face make different contributions to recognition, we assign weights to regions for weighted voting. We construct a decision dictionary according to the recognition results of selected regions in the training phase, and this dictionary is used in a self-defined loss function to obtain weights. The final identity of test sample is the weighted voting of selected regions. In this paper, we combine the WVDR strategy with CRC and SRC separately, and extensive experiments show that our method outperforms the baseline and some representative algorithms.

***key words:*** *discriminative regions, small sample size, occlusion, weighted strategy, face recognition*

## 1. Introduction

Face recognition is one of the most popular and challenging problems in computer vision. Many representative methods, such as SRC [1] and CRC [2], have achieved good results in the controlled condition. However, face recognition with occlusion or small training size is still challenging.

Wright *et al.* [1] first apply the Sparse Representation based Classification (SRC) for face recognition (FR). Zhang *et al.* [2] propose Collaborative Representation based Classification (CRC) and claim that it is the CR instead of the $l_1$-norm sparsity that truly improves the FR performance. However, the performance of classifiers (e.g. SVM [3], SRC and CRC) declines dramatically if the training sample size is small. Some works have been done to tackle the Small Sample Size (SSS) problem. The Extended SRC [4] algorithm constructs an auxiliary intra-class variant dictionary to represent the variations between training and test images, while the construction of the dictionary needs extra data. Patch-based methods are another effective way to solve the SSS problem. In [5], Zhu *et al.* propose the patch-based CRC and multi-scale ensemble. Gao *et al.* [6] propose the Regularized Patch-based Representation to solve the SSS problem. However, patch-based methods are sensitive to the patch size [7], and haven't noticed the texture distribution of a face image.

Images with disguise or occlusion are hard to classify. The recognition rate of many classifiers (e.g. SVM and SRC) decreases rapidly when images occluded. Local Contourlet Combined Patterns (LCCP) [8] reports a good performance in non-occlusion images but the recognition rate decreases in occlusion condition. There are some improvements [9], [10] for occlusion problem. The recent probabilistic collaborative representation (ProCRC) [10] jointly maximizes the likelihood of test samples with multiple classes.

Instead of splitting the image into patches of same size, we extract the face regions according to an alignment algorithm [11]. Some regions, such as eyes and nose, are discriminative for recognition. In addition, different regions have different representation abilities. As Fig. 1 shows, discriminative ability of regions is affected by type of region and training size. So it's reasonable that the regions are assigned with different weights.

In this paper, we propose a method termed Weighted Voting of Discriminative Regions (WVDR), in which, discriminative regions are extracted from face images and

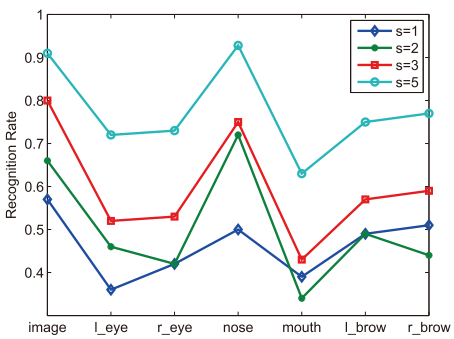


**Fig. 1** Recognition rates (AR database) when using only a single region. The *s* represents the number of training samples per person. The *X*-axis represents the regions extracted from face, and the *image* means the whole face image.

Manuscript received June 5, 2017.
Manuscript revised July 16, 2017.
Manuscript publicized August 4, 2017.

[†]The authors are with Shenzhen Key Lab. of Information Sci & Tech, Shenzhen Engineering Lab. of IS & DCP Department of Electronic Engineering, Graduate School at Shenzhen, Tsinghua University, China.

*This work was supported in part by the Natural Science Foundation of China under Grant No.61471216, in part by the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under Grant No. JCYJ20150831192224146, No. JCYJ20150601165744635 and No. JCYJ20150331151358138, and in part by Open Foundation from Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering.

a) E-mail: grq15@mails.tsinghua.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2017EDL8124

weights are learned from a decision dictionary in training phase. The decision dictionary contains recognition information of each region, and the weights computation is fast since it can be obtained from closed form solution. In the test phase, the weights achieved from the training phase are applied to regions, and the final predicted label is achieved by weighted voting of discriminative regions. The WVDR strategy not only performances well in the Small Sample Size (SSS) problem, but also is robust to occlusions.

The rest of the paper is organized as follows. In Sect. 2 we describe our model in detail. We conduct extensive experiments to test our model in Sect. 3 and summarize the work in Sect. 4.

## 2. Weighted Voting of Discriminative Regions

In this section, we describe our algorithm in detail. The framework of our method is illustrated in Fig. 2. Firstly, we briefly introduce the baseline classifier (SRC and CRC). Then, we explain why the discriminative regions is suitable in our algorithm. Finally, we introduce the Weighted Voting of Discriminative Regions (WVDR).

### 2.1 SRC and CRC

Denote by $X_i$ the data set of $i$-th class, and each column of $X_i$ is a sample from $i$-th class. Suppose we have $k$ classes and training set is $X$, $X = [X_1, X_2, \ldots, X_k]$. Given a test sample $y$, usually $y \approx X\alpha$ holds well. In SRC [1], the coding vector $\alpha$ is described as follows:

$$\hat{\alpha} = argmin_\alpha \left\{ \|y - X\alpha\|_2^2 + \lambda\|\alpha\|_1 \right\} \tag{1}$$

CRC [2] is a little different from SRC: using the $l_2$-norm rather than $l_1$-norm. The coding vector is obtained by:

$$\hat{\alpha} = argmin_\alpha \left\{ \|y - X\alpha\|_2^2 + \lambda\|\alpha\|_2 \right\} \tag{2}$$

The recognition of $y$ is $identity(y) = \arg\min_i(r_i(y))$. Where $r_i(y) = \|y - X_i\hat{\alpha}_i\|^2/\|\hat{\alpha}_i\|^2$, and $\alpha_i$ is the coefficient related to $i$-th class.

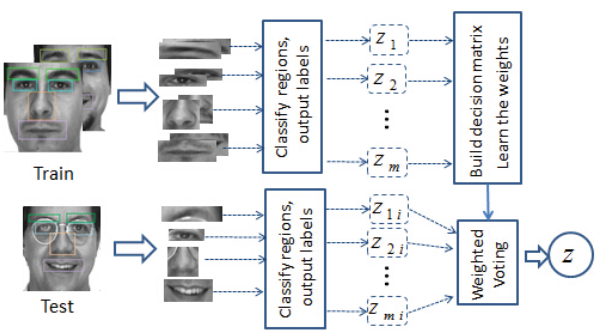


**Fig. 2** The framework of our method. Every region outputs a label by classification. In the training phase, we obtain the weights of regions. Then we combine all the outputs as a final result based on Weighted Voting for the test image.

### 2.2 Why Using Discriminative Regions?

There are mainly three reasons to utilize the Discriminative Regions. First, when the available training samples are limited, classification methods (e.g. SRC, CRC) cannot get a satisfactory result. Extracting discriminative regions, like patches in PCRC [5], can solve the problem of sample-limited well. Second, discriminative parts, like eyes and nose, possess abundant texture, which make greater contribution to recognition [12], [13]. Additionally, voting of multi regions is robust to occlusion. As the second row of Fig. 3 shows, the original image is severely affected, but the selected regions remain impervious or only partly affected. The regions are classified independently, so even when the discriminative region (such as left eye) is occluded, the other parts can vote for the final result normally. Experiments in Sect. 3 demonstrate the robustness of our strategy when occlusion size is not enormous. However, if two or more regions is occluded, the WVDR strategy becomes unstable.

### 2.3 Weighted Voting of Multi-Regions

Though we extract the discriminative regions in the face image, the effect of different regions varies a lot with the change of training size. The accuracy in Fig. 1 indicates that regions receive different recognition rates, and the recognition rate is affected by the kind of regions and training size.

We aim to assign different weights to different regions. In this section, we learn the weights by minimizing a loss function, and the loss function reflects the errors of training phase.

As Fig. 2 shows, 6 regions are selected. The voting process contains 7 items: the 6 regions and the whole image. Every item outputs a label $\hat{y}_i^r$ by classifier (e.g. CRC) in training phase. Given a set of samples $S$, $x_i \in S$, $i = 1, 2, \ldots, n$. The recognition result of region $r$ in image $x_i$ is $\hat{y}_i^r$. In order to measure the discrimination ability of regions, we construct a decision dictionary. Decision dictionary $D = \{d_i^r\}^{n\times m}$, $i = 1, 2, \ldots, n$, $r = 1, 2, \ldots, m$, is obtained as follows:

$$d_i^r = \begin{cases} 1 & (\hat{y}_i^r == l_i) \\ -1 & otherwise \end{cases} \tag{3}$$

where $m$ is the number of regions, and $n$ is the number of

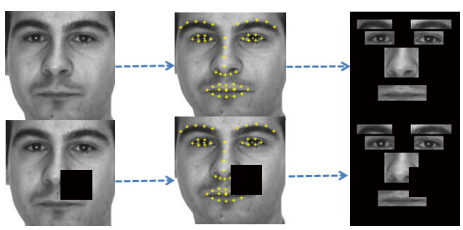

**Fig. 3** Extracting discriminative regions. We align the face and then extract the regions around the facial points. The second row indicates the situation of occlusion.

Please confirm.

samples in $S$. When recognition of sample $i$ in region $r$ is right, we set $d_i^r = 1$. $d_i^r = -1$ for the wrong recognition.

The loss of sample $y_i$ is defined as:

$$\xi(y_i) = \left(\sum_{r=1}^{m} w^r (1 - d_i^r)\right)^2 \tag{4}$$

where $w^r$ is the weight assigned to region $r$. In the training phase, we minimize the training error. For the whole sample set $S$, the loss function is defined as follows:

$$\begin{aligned} L(S) &= \sum_{i=1}^{n} \xi(y_i) \\ &= \sum_{i=1}^{n} \left(\sum_{r=1}^{m} w^r (1 - d_i^r)\right)^2 = \|e_1 - Dw\|^2 \end{aligned} \tag{5}$$

where $w = \left[w^1, w^2, \ldots, w^m\right]^T$, $\sum_{r=1}^{m} w^r = 1$, $w^r > 0$. $e_1$ is a column vector whose elements are all *one* and length is $n$.

In CRC, $l_2$-norm is used to avoid over fitting. In WVDR we use the $l_2$-norm enhance the more discriminative regions, which reflects in the second item of Eq. (6). It is reasonable to use the $l_2$-norm rather than $l_1$-norm, because we want all the selected regions involved. Just as Fig. 3 shows, though one or two regions are occluded,the recognition is still robust.

As for the restriction $\sum_{r=1}^{m} w^r = 1$, we rewrite it as $e_2^T w = 1$, where $e_2$ is a column vector whose elements are all *one* and length is $m$. The obtained weight vector is written as:

$$\hat{w} = \arg\min_{\hat{w}} \left\{ \|e_1 - Dw\|^2 + \lambda \|w\|^2 + \gamma \left(e_2^T w - 1\right)^2 \right\} \tag{6}$$

Because $e_2^T w e_2 = e_2 e_2^T w$, the close form solution of Eq. (6) is

$$\hat{w} = \left(D^T D + \lambda I + \gamma e_2 e_2^T\right)^{-1} \left(\gamma e_2 + D^T e_1\right) \tag{7}$$

where $e_2 e_2^T = \{1\}^{m \times m}$ is the matrix whose elements are all *one*, and $I$ is the identity matrix.

As Fig. 2 shows, the final output label is obtained by regions voting. And the importance of each region is reflected on the weights.

## 3. Experiment

We verify the effectiveness of our model on the AR [14], EYB [15] and LFW [16] databases. VDR represents the WVDR when all the weights are the same. We run the algorithms 10 times in AR and 15 times in EYB and LFW, and *average* ± *standard deviation* is reported. The results with asterisk are taken from reference [5], and the other results are obtained by running public codes or our algorithm.

When extracting the regions from image, we resize the face to $140 \times 140$. While implementing experiments on the whole image, we resize the image of AR to $60 \times 43$ and the image of EYB and LFW to $32 \times 32$.

We conduct regions of different sizes for comparison, and the results show that the recognition rate is not sensitive to region size. Also, we assigns different values to $\lambda$ and $\gamma$, and the accuracy varies little. Thus, it's reasonable that we select only one representative size and set $\lambda = \gamma = 0.01$ to do the following work. The region size we choose is: brow $37 \times 21$, eye $33 \times 21$, nose $41 \times 41$, mouth $71 \times 29$. The center of regions is determined by face alignment [11].

We change the value of $S$ (training samples per person) in the AR database, and compare the results of our method with the representative algorithms. The results are demonstrated in Table 1. We artificially add some contiguous occlusions and the positions are randomly generated (illustrated in second row of Fig. 3). Experimental results of different scale occlusions are presented in Table 2 ($10 \times 10$ represents that $10 \times 10$ occlusions are added to test images), which prove that our method is robust to occlusions.

EYB contains 38 people under 9 poses and 64 different illumination conditions. LFW [16] is a challenging database which is widely used in face verification [12], [17]. Instead, we select the individuals with no less than 10 images from LFW to do the face recognition task.

In Table 3, we compare our method with representative methods on EYB and LFW. All the experiments under the

**Table 1** The accuracy on AR.

| S | 1 | 2 | 5 |
|---|---|---|---|
| CRC [2] | 42.9±14.6* | 69.9±12.6* | 89.1±6.2* |
| SRC [1] | 44.9±14.8* | 69.7±14.8* | 88.2±5.7* |
| PCRC [5] | 65.4±20.9* | 82.2±11.3* | 92.9±6.7* |
| LCCP [8] | 63.5±15.6 | 79.8±13.7 | 94.1±8.3 |
| ProCRC [10] | 67.6± 15.3 | 85.8±14.1 | 95.2 ± 8.7 |
| VDR+SRC | 66.4±14.2 | 83.4±13.2 | 95.0±10.3 |
| VDR+CRC | 66.1±13.6 | 84.0±14.4 | 95.5±9.1 |
| **WVDR+SRC** | **73.4±12.4** | 85.3±11.3 | 96.5±8.9 |
| **WVDR+CRC** | 72.7±14.8 | **87.4±12.6** | **97.1±9.2** |

**Table 2** The accuracy on AR (with occlusion).

| Occlusion | $10 \times 10$ | $15 \times 15$ | $20 \times 20$ |
|---|---|---|---|
| CRC [2] | 60.0±8.9 | 47.3±9.5 | 34.2±11.7 |
| SRC [1] | 66.5±7.2 | 52.1±10.1 | 41.5±12.0 |
| PCRC [5] | 79.8±10.2 | 76.0±9.2 | 67.9±9.8 |
| LCCP [8] | 82.5±11.4 | 76.3±12.9 | 65.5±15.4 |
| ProCRC [10] | 92.6±8.4 | **91.5± 9.5** | **86.6±11.6** |
| VDR+SRC | 85.4±6.4 | 81.3±5.2 | 72.4±5.8 |
| VDR+CRC | 84.7±4.9 | 83.5±5.9 | 71.2±6.0 |
| **WVDR+SRC** | **93.5±8.5** | 90.6±7.8 | 83.7±8.6 |
| **WVDR+CRC** | 93.2±7.6 | 91.3±8.3 | 85.4±10.5 |

**Table 3** The accuracy on EYB and LFW (S = 5).

| Method | EYB | LFW | year |
|---|---|---|---|
| SRC [1] | 89.0±12.5* | 44.1±2.6* | 2009 |
| CRC [2] | 87.8±13.7* | 42.0±3.2* | 2011 |
| PCRC [5] | 92.0±8.2* | 42.9±2.6* | 2012 |
| LCCP [8] | 91.4±15.3 | 44.5±3.5 | 2016 |
| ProCRC [10] | 94.4±10.6 | 45.2±3.3 | 2016 |
| VDR+SRC | 93.2±8.5 | 47.0±4.1 | - |
| VDR+CRC | 93.6±9.3 | 46.3±3.8 | - |
| **WVDR+SRC** | 94.7±13.5 | 48.4±4.6 | - |
| **WVDR+CRC** | **95.6±11.2** | **49.7±3.1** | - |

same setting. Our method is competitive, especially in LFW.

The WVDR strategy can greatly improve the performance of classification methods, especially in small sample size problems (e.g. CRC: from 89.1% to 97.1% in AR, SRC: from 89.0 % to 94.7% in EYB). The weighted strategy is reasonable because if we don't assign different weights for voting (in VDR), the performance improvement is limited (showed in Table 1–3). When the images are occluded, WVDR strategy not only beats the baseline methods (SRC and CRC) by a essential margin but also outperforms some representative methods (e.g. PCRC and LCCP). However, WVDR strategy is a little weaker than ProCRC when occlusion size is large.

## 4. Conclusion

This paper presents a novel strategy named Weighted Voting of Discriminative Regions for robust face recognition. Our method extracts the discriminative regions based on face alignment, which has been validated to be more efficient than patch-based methods. Then, the learned weight is applied on regions, which makes the voting more robust. We conducted extensive experiments on 3 databases. The results show that our method is competitive in both the Small Sample Size problems and the occlusion conditions. Our future works would include improving FR performance when the occlusion size is enormous.

### References

[1] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.2, pp.210–227, 2009.

[2] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," 2011 IEEE International Conference on Computer Vision (ICCV), pp.471–478, 2011.

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

[4] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.9, pp.1864–1870, 2012.

[5] P. Zhu, L. Zhang, Q. Hu, and S.C.K. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," Computer Vision – ECCV 2012, Lecture Notes in Computer Science, vol.7572, pp.822–835, Springer, 2012.

[6] 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.

[7] S. Chen, J. Liu, and Z.-H. Zhou, "Making FLDA applicable to face recognition with one sample per person," Pattern Recognit., vol.37, no.7, pp.1553–1555, 2004.

[8] Y. Wang, S. Yu, W. Li, L. Wang, and Q. Liao, "Face recognition with local contourlet combined patterns," IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1273–1277, 2016.

[9] R. Min, A. Hadid, and J.-L. Dugelay, "Improving the recognition of faces occluded by facial accessories," 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), pp.442–447, 2011.

[10] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2950–2959, 2016.

[11] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp.3444–3451, 2013.

[12] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.1891–1898, 2014.

[13] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," IEEE Trans. Image Process., vol.18, no.8, pp.1885–1896, 2009.

[14] A.M. Martinez, "The ar face database," CVC Technical Report, vol.24, 1998.

[15] 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," IEEE Trans. Pattern Anal. Mach, Intell., vol.23, no.6, pp.643–660, 2001.

[16] G.B. Huang and E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep, pp.14–003, 2014.

[17] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2892–2900, 2015.

AR