A New Feature Fusion Approach Based on LBP and Sparse Representation and Its Application to Face Recognition

He-Feng Yin and Xiao-Jun Wu

School of IoT Engineering, Jiangnan University, Wuxi 214122, China
yinhefeng@126.com, wu_xiaojun@yahoo.com.cn

Abstract. In this paper, we propose a new feature fusion approach based on local binary pattern (LBP) and sparse representation (SR). Firstly, local features are extracted by LBP and global features are sparse coefficients which are obtained via decomposing samples based on the over-complete dictionary. Then the global and local features are fused in a serial fashion. Afterwards PCA is used to reduce the dimensionality of the fused vector. Finally, SVM is employed as a classifier on the reduced feature space for classification. Experimental results obtained on publicly available databases show that the proposed feature fusion method is more effective than other methods like LBP+PCA, Gabor+PCA and Gabor+SR in terms of recognition accuracy.

Keywords: Feature fusion, local binary pattern, sparse representation, support vector machine, face recognition.

1 Introduction

Automatic face recognition [1] remains one of the most visible and challenging research topics in computer vision, machine learning and biometrics. It is widely applied to different fields including biometric authentication, security applications and human computer interaction. Compared with other biometrics, such as fingerprint identification and palm identification, face recognition has the advantages of being convenient, immediate and well accepted.

The question of which low-dimensional features of an object image are the most relevant or informative for classification is a central issue in face recognition. Conventional facial features can be roughly divided into global features (PCA [2], LDA [3], LPP [4], etc.) and local features (LBP [5], SIFT [6], etc.). However, both the global and local features are not rich enough to capture all of the classification information available in the image, in addition, researches have shown that different features have different classification capabilities and a fusion scheme that harnesses various features is likely to improve the overall performance.

There are three levels of information fusion, i.e. pixel level, feature level and decision level. The decision level fusion, represented by multi-classifier combination, has been one of the hot research topics on pattern recognition [7-10]. In recent years,

Z.-H. Zhou, F. Roli, and J. Kittler (Eds.): MCS 2013, LNCS 7872, pp. 364–373, 2013.

some feature level fusion methods have been proposed, for instance, Sun et al. [11] proposed a novel feature fusion method. Firstly, two groups of feature vectors are extracted with the same pattern, then a correlation criterion function is established between the two groups of feature vectors, finally their canonical correlation features are extracted to form effective discriminant vectors for recognition. Huang [12] put forward an efficient face representation and recognition method, which combines the both information between rows and those between columns from two-directional 2DPCA on fusion face image and the optimal discriminative information from column-directional 2DLDA. Song [13] provided a method based on the feature fusion of the local and global features, local features are extracted from sub-images and global features are obtained via PCA. Chowdhury et al. [14] presented a fusion method, first of all, face images are divided into a number of non-overlapping sub-images, the G-2DFLD method is applied to each of these sub-images as well as to the whole image to extract local as well as global discriminant features respectively. These extracted local and global features are fused to form a large feature vector and FLD method is applied on it to reduce its dimensionality. Nevertheless, the above fusion methods are largely dependent on the dimensionality of features, and in low-dimensional feature space, recognition accuracy of these methods is not that high.

However, within the framework of sparse representation, the precise choice of feature space is no longer critical. What is crucial is that the dimensionality of the feature space is sufficiently large and that the sparse representation is correctly computed [15]. In addition, according to related researches about local binary pattern (LBP), features coded by LBP have highly discriminative power [16], this property makes it suitable for image classification tasks. Inspired by these findings, we intend to use the fused features of sparse coefficient and local features extracted by LBP to improve the recognition performance.

The remainder of this paper is organized as follows: LBP and sparse representation are reviewed in Section 2 and Section 3 respectively. Section 4 presents the proposed method. Experiments are conducted on publicly available databases to verify the effectiveness of the proposed method in Section 5. Finally, conclusions are drawn in Section 6.

2 Local Binary Pattern

The LBP operator was first introduced by Ojala [17] and used as texture descriptor. Then Ahonen [5] applied it to face recognition and obtained outstanding results, which demonstrates that LBP is able to well describe face images.

The original LBP operator was defined as a window of size 3×3. This operator uses the value of the center pixel as a threshold, and the 8 surrounding pixels whose value is higher than or equal to the value of the threshold is assigned a binary value 1, otherwise the value is 0. When this process is accomplished, 8 values can be read start from the top left corner in the clockwise direction. The 8-bit binary number or its equivalent decimal number can be assigned to the center pixel and it can describe the texture information of an image. The basic LBP operator is illustrated in Fig. 1.

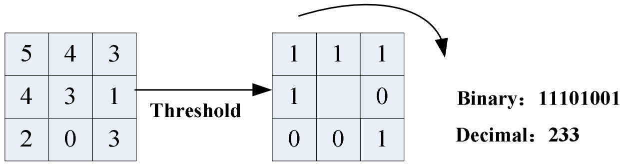

Fig. 1. The original LBP operator

In order to facilitate the analysis of textures with different scales, the basic LBP operator is extended by combining neighborhoods with different radius. In this case, P points on the edge of a circle, whose radius is R, are sampled and compared with the value of the center pixel. For ease of presentation, the notation (P,R) is employed to formulate P sampling points on a circle of radius of R. See Fig. 2 for an example of circular neighborhoods.

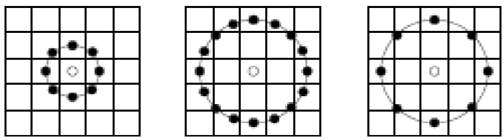

Fig. 2. The circular (8,1), (16,2) and (8,2) neighborhoods

Another extension of the original LBP operator is the definition of so called uniform patterns. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular [18]. Experimental results have demonstrated uniform patterns can describe most of the texture information, at the same time, they have strong ability to do classification tasks.

Generally, when we extract features from face images, we can divide the face image into small blocks. And features are extracted from each block independently. The descriptors are then concatenated to form a global description of the face image. In this way we can obtain a description of the face image on local and holistic levels. In this paper, uniform patterns of (8,1) are applied to extracted LBP features.

3 Sparse Representation (SR)

Theoretical results show that well-aligned images of a convex, Lambertian object lie near a low-dimensional feature space of the high-dimensional image space [19]. This is the only prior knowledge about the training samples in SR. The idea of SR is presented as follows [15].

Suppose we have C distinct classes, given sufficient training samples of the i-th object class, the size of face images is $w \times h$, and the total number of samples of i-th class is n_i. We stack the n_i training images from the i-th class as columns of a matrix $A_i = [v_{i,1}, \ldots, v_{i,n_i}] \in R^{m \times n_i}$ (m=w×h). For a test sample $y \in R^m$ belongs to this class, according to linear subspace theory, y can be approximated by the linear combination of the samples within A_i, i.e.

$$y \approx \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \cdots + \alpha_{i,n_i} v_{i,n_i} \tag{1}$$

$\alpha_{i,j} \in R, j = 1, 2, \cdots, n_i$.

Since the initial identity of the test sample y is unknown, let A be the concatenation of the n training samples from all the C classes, where $\sum_{i=1}^{C} n_i = n$, then we can define a new matrix A:

$$\begin{aligned} A &= [A_1, A_2, \cdots, A_C] \\ &= [v_{1,1}, \ldots, v_{1,n_1}, \ldots, v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}, \ldots, v_{C,1,\ldots,} v_{C,n_C}] \end{aligned} \tag{2}$$

If we use the new matrix A to represent the test image y, that is

$$y = Ax_0 \in R^m \tag{3}$$

where $x_0 = [0, \ldots, 0, \ldots, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, \ldots, 0, \ldots, 0]^T \in R^n$ is a coefficient vector whose entries are zero except those associated with the i-th class, and A is referred to as dictionary.

In robust face recognition, the system $y = Ax$ is always ill-determined, so its solution is not unique, but we just need to find a locally optimal solution. Conventionally, this problem is settled by choosing the minimum l^2-norm solution. However, the solution is non-sparse and it has no discriminative information. This motivates us to seek the sparsest solution to $y = Ax$, leading to the following optimization problem:

$$(l^0)\ x_0 = \arg\ \min\ \| x \|_0,\ subject\ \ to\ Ax = y \tag{4}$$

where $\|\cdot\|_0$ denotes the l^0-norm, which counts the number of nonzero elements in a vector.

However, the problem of finding the sparsest solution of an ill-determined system of linear equations is NP-hard. Recent progress in the theory of sparse representation and compressed sensing reveals that if the solution x_0 is sparse enough, the solution to the l^0-minimization problem (4) is equal to the following l^1-minimization problem [20]:

$$(l^1) x_1 = \arg\ \min\ \| x \|_1, subject\ \ to\ Ax = y \tag{5}$$

To solve the l^1-minimization problem, one can use gradient projection method [21], homotopy algorithm [22], iterative shrinkage-thresholding [23] etc.

In order to guarantee the coefficient vector x has the form $[0, \ldots, 0, \alpha, 0, \ldots, 0]$ where all the non-zero entries are together, we solve this optimization problem:

$$\min_x \| y - Ax \|_2 + \lambda_1 \| x \|_2^2 + \lambda_2 \| x \|_1 \tag{6}$$

The l_1 penalty in the above expression promotes sparsity of the coefficient vector x, while the quadratic l_2 penalty encourages grouping effect, i.e. selection of a group of correlated training samples.

4 Proposed Feature Fusion Method

Wavelet transform has been introduced in our method to perform the preprocessing of the face images, it can reduce noise of images, and the low frequency component is a coarser approximation to the original image. Thus the wavelet image should be more suitable for recognition.

Given all that, the procedure of the proposed method is presented as follows:

1. Perform wavelet transform to the original image and obtain its 1-level low-frequency component *L*.
2. Divide the 1-level low frequency component into small blocks, then extract LBP features for each small block.
3. Concatenate the LBP features of all the small blocks to form the local feature of the original image.
4. Based on the over-complete dictionary (which contains all the training samples), the same original image can be decomposed to obtain its sparse coefficient, i.e., the global feature.
5. Then the local and global features are fused in a serial fashion [24], after the dimensionality of the fused feature is reduced, it can be used for recognition.

Framework of the proposed method and other methods that will be compared with in this paper is depicted in Fig. 3.

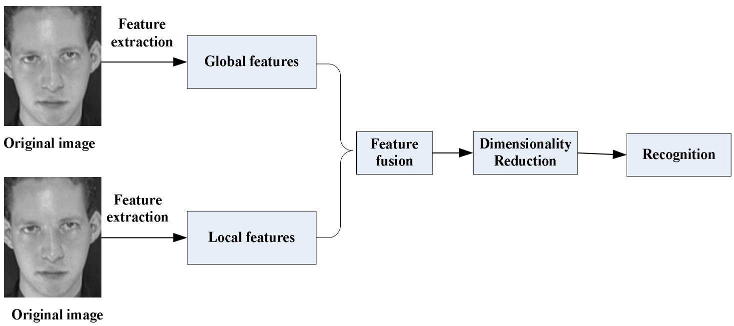

Fig. 3. Framework of the methods considered in this paper

5 Experiments and Analysis

In this section, we conduct experiments on publicly available databases for face recognition. The ORL and XM2VTS databases are used to verify the performance of the proposed method and its competing methods:

PCA: global features extract by PCA.
LBP: local features extract by LBP.
SR: sparse representation of the sample, i.e. sparse coefficient.
Gabor+PCA: fused features extracted by Gabor filter and PCA.
Gabor+SR: fused features extracted by Gabor filter and sparse coefficient.
LBP+PCA: fused features extracted by LBP and PCA.

When extracting local features based on LBP, the original face image is preprocessed by wavelet transform. In this experiment, the basis function of wavelet transform is *coif4*. In SR, the error tolerance ε is 0.05. We use Gabor filter at five different scales and eight orientations, thus we obtain 40 Gabor filters. The global and local features are fused in a serial fashion. Then PCA is utilized to do dimensionality reduction. Finally, linear SVM is employed for classification and the strategy for multi-class classification is one-against-one approach [25].

5.1 Experiments on the ORL Database

The ORL database contains images from 40 individuals, each providing 10 different images. For some subjects, the images were taken at different times. The facial expressions (open or closed eyes, smiling or non-smiling) and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent. All images are gray-scale and have a resolution of 92×112 pixels. Half of the images per subject are chosen as training samples, the reminder for testing, and the face image is divided into 4×4 blocks when extracting the LBP features. Fig. 4 shows the recognition performance for various methods, in conjunction with different feature dimensionality. Table 1 shows the detailed recognition accuracy of the methods considered and Table 2 records the computation time of Gabor+SR and the proposed method.

Table 1. Recognition rate(%) of different methods on the ORL database and the associated dimensionality of feature

Dimensionality	10	30	50	70	90
PCA	93.5%	92%	94.5%	93.5%	91.5%
LBP	83.5%	93.5%	95.5%	97%	97%
SR	82%	94%	93.5%	93.5%	93%
Gabor+PCA	92%	94%	95%	95.5%	95%
Gabor+SR	82%	96%	96.5%	97.5%	98%
LBP+PCA	92%	94%	95%	95.5%	95%
Proposed	97%	97%	97%	97%	97%

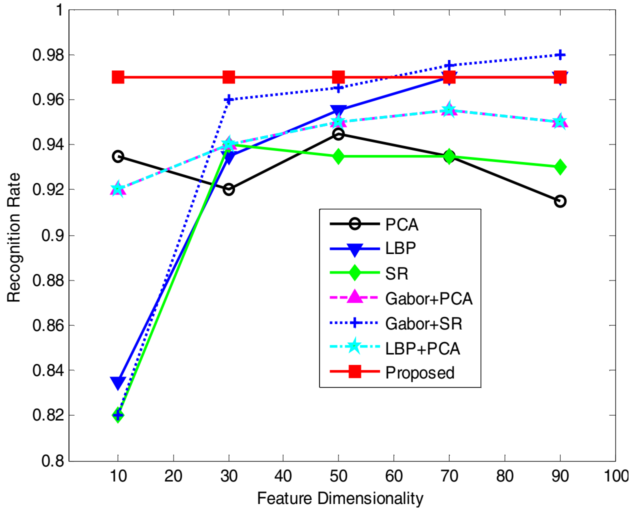

Fig. 4. Curves of recognition rate by different methods versus feature dimensionality on the ORL database

Table 2. Computation time(s) of Gabor+SR and the proposed method on the ORL database and the associated dimensionality of feature

Dimensionality	10	30	50	70	90
Gabor+SR	115.43s	116.03s	116.66s	117.42s	118.04s
Proposed	16.37s	16.42s	16.47s	16.55s	16.61s

5.2 Experiments on the XM2VTS Database

The XM2VTS database is a multi-modal database which consists of video sequences of talking faces recorded for 295 subjects at one month intervals. The data has been recorded in 4 sessions with 2 shots taken per session. From each session two facial images have been extracted to create an experimental face database of size 55×51. In our experiment, we chose a subset of the dataset consisting of 100 subjects. For each subject, four images are used as training samples, the rest for testing, and the face image is divided into 8×8 blocks when extracting the LBP features. The comparison of competing methods is given in Fig. 5 and Table 3. Computation time of Gabor+SR and the proposed method is recorded in Table 4.

Table 3. Recognition rate(%) of different methods on the XM2VTS database and the associated dimensionality of feature

Dimensionality	5	10	20	30	35
PCA	33.5%	58.75%	78.25%	86.5%	87.75%
LBP	35.75%	58.25%	81%	87.25%	88.5%
SR	27%	62.5%	80.5%	85.75%	88%
Gabor+PCA	45%	69%	80.5%	83.25%	83.75%
Gabor+SR	52.75%	79.75%	92%	95%	94.75%
LBP+PCA	45%	68.75%	80.25%	83.25%	83.75%
Proposed	96%	96%	96%	96%	96%

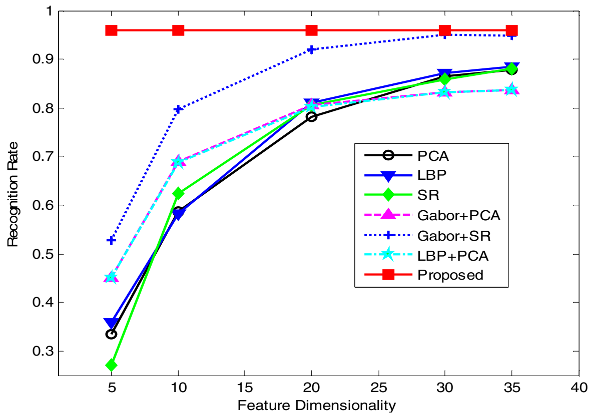

Fig. 5. Curves of recognition rate by different methods versus feature dimensionality on the XM2VTS database

Table 4. Computation time(s) of Gabor+SR and the proposed method on the XM2VTS database and the associated dimensionality of feature

Dimensionality	5	10	20	30	35
Gabor+SR	470.45s	470.61s	471.80s	473.14s	473.73s
Proposed	72.38s	72.66s	73.22s	73.77s	74.15s

Based on the above experimental results obtained on ORL and XM2VTS databases, we have the following observations:

1. As feature dimensionality increases, performance of LBP is better than that of PCA, this indicates that local features may contain more discriminative information.
2. When we fuse global features (e.g. features extracted by SR) with local features (e.g. Gabor features), performance of global features is boosted. This demonstrates that fused features can improve the overall performance.

3. By and large, the proposed method is more competitive than other methods, not only the performance of the proposed method remains stable, but the computation time is acceptable. Though performance of Gabor+SR is better than that of LBP+SR on ORL database, it is computationally expensive, and its computation time is about 7 times that of our method.

6 Conclusions

In this paper, we propose a new feature fusion approach based on LBP and sparse representation. Firstly, local features are extracted by LBP and global features are sparse coefficients which are obtained via decomposing samples based on the over-complete dictionary. Then the global and local features are fused in a serial fashion. Experiments conducted on the ORL and XM2VTS databases show the feasibility and effectiveness of the new method. However, in this paper, we do not explore other feature fusion methods, so in future, we will investigate other methods and come up with a better approach for robust face recognition.

References

1. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and Machine Recognition of Faces: A Survey. Proceedings of the IEEE 83(5), 705–741 (1995)
2. Turk, M., Pentland, A.: Face Recognition Using Eigenfaces. In: CVPR, pp. 586–591 (1991)
3. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE PAMI 9(7), 711–720 (1997)
4. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face Recognition Using Laplacianfaces. IEEE Trans. Pattern Analysis and Machine Intelligence 27(3), 328–340 (2005)
5. Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
6. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the Use of SIFT Features for Face Authentication. In: Proc. of IEEE Conf. on Biometrics, in Association with CVPR Biometrics, p. 35 (2006)
7. Huang, Y.S., Suen, C.Y.: Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals. IEEE Trans. Pattern Anal. Mach. Intell. 7(1), 90–94 (1995)
8. Constantinidis, A.S., Fairhurst, M.C., Rahman, A.F.R.: A New Multi-Expert Decision Combination Algorithm and Its Application to the Detection of Circumscribed Masses in Digital Mammograms. Pattern Recognition 34(8), 1528–1537 (2001)
9. Jin, X.-Y., Zhang, D., Yang, J.-Y.: Face Recognition Based on a Group Decision-Making Combination Approach. Pattern Recognition 36(7), 1675–1678 (2003)
10. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers. Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)

11. Sun, Q.-S., Zeng, S.-G., Liu, Y., Heng, P.-A., Xia, D.-S.: A New Method of Feature Fusion and Its Application in Image Recognition. Pattern Recognition 38(12), 2437–2448 (2005)
12. Huang, G.H.: Fusion $(2D)^2$PCALDA: A New Method for Face Recognition. Applied Mathematics and Computation 216(11), 3195–3199 (2010)
13. Song, L.: Face Recognition Based on Feature Fusion. In: Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC), pp. 1524–1527 (2011)
14. Chowdhury, S., Sing, J.K., Basu, D.K., Nasipuri, M.: Face Recognition by Fusing Local and Global Discriminant Features. In: 2nd International Conference on Emerging Applications of Information Technology (EAIT), pp. 102–105 (2011)
15. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust Face Recognition via Sparse Representation. IEEE PAMI 31(2), 210–227 (2009)
16. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
17. Ojala, T., Pietikinen, M.: A Comparative Study of Texture Measures with Classification Based on Feature Distribution. Pattern Recognition 29(1), 51–59 (1996)
18. Ahonen, T., Hadid, A., Pietikainen, M.: Face Description with Local Binary Patterns: Application to Face Recognition. PAMI 28(12), 2037–2041 (2006)
19. Basri, R., Jacobs, D.: Lambertian Reflectance and Linear Subspaces. IEEE Trans. Pattern Analysis and Machine Intelligence 25(2), 218–233 (2003)
20. Donoho, D.: For Most Large Underdetermined Systems of Linear Equations the Minimal l^1-Norm Solution Is Also the Sparsest Solution. Comm. Pure and Applied Math. 59(6), 797–829 (2006)
21. Figueiredo, M., Nowak, R., Wright, S.: Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. IEEE Journal of Selected Topics in Signal Processing 1(4), 586–597 (2007)
22. Drori, I., Donoho, D.: Solution of l^1-Minimization Problems by LARS/Homotopy Methods. In: ICASSP, pp. 636–639 (2006)
23. Hale, E., Yin, W., Zhang, Y.: A Fixed-Point Continuation Method for l^1-Regularized Minimization with Applications to Compressed Sensing. Technical Report, Rice University (2007)
24. Yang, J., Yang, J.-Y., Zhang, D., Lu, J.: Feature Fusion: Parallel Strategy vs. Serial Strategy. Pattern Recognition 36(6), 1369–1381 (2003)
25. Guo, G.-D., Li, S.-Z., Chan, K.-L.: Support Vector Machines for Face Recognition. Image and Vision Computing 19(9-10), 631–638 (2001)