

## Face Recognition Using Plurality Voting and Collaborative Representation based on the Bit-plane Information

Dongmei Wei<sup>1,2</sup> and Weidong Zhou

<sup>1</sup>*School of Information Science and Engineering, Shandong University  
27 Shanda Nan Road, Jinan, China 250100*

<sup>2</sup>*College of Physics and Electronics, Shandong Normal University  
88 Wenhua Dong Road, Jinan, China 250014*

*School of Information Science and Engineering, Shandong University  
27 Shanda Nan Road, Jinan, China 250100  
weidongmei2@163.com, wdzhou@sdu.edu.cn*

### Abstract

*As a new representation technique, collaboration representation is used to represent the test sample with all training samples from all classes. Plurality voting is one of the most widely used combination strategies in pattern recognition. This paper presents a new and an efficient face recognition approach using plurality voting and collaboration representation based on the binary bit-plane images. First, face gray images are equalized and decomposed. Next, employing collaboration representation and the corresponding five bit-plane images which have more discrimination information, five identities about the same test image can be obtained. Finally, these five identities vote to the true identity of the test image. If the plurality voting fails, the true identity of the test image will be decided by virtual weight sum face images constructed by 8 bit-plane images with collaboration representation. Weight vector, which is important for the virtual images, is determined by the recognition rate and the order of bit planes. The extensive experiments demonstrate the proposed approach has the higher right recognition accuracy and its speed is faster than the SRC prominently.*

**Keywords:** *Bit-plane, histogram equalization, plurality voting, virtual weight sum image, collaborative representation*

### 1. Introduction

Face recognition has become a hot research topic in pattern recognition and computer vision due to its broad applications in human-computer interaction, security, law enforcement, and entertainment. However, since the face is not rigid, the performance of the face recognition approach is affected by expression, illumination, occlusion, pose, age and so on. Hence, face recognition is still a hard problem presently.

Face recognition has been studied widely and lots of approaches have been proposed. Principal Component Analysis (PCA) [1] and Linear Discriminate Analysis (LDA) [2] have been used widely as subspace learning techniques. PCA projects the high-dimensional face images into the low-dimensional feature space along the directions of maximal variances and preserves the Euclidean distances between face images. As a supervised algorithm, LDA finds the optimal discrimination directions and projects the images into the linear subspace. Some researches have shown the face images possibly reside on a nonlinear submanifold, while both PCA and LDA fail to discover the underlying structure. Locality Preserving Projections (LPP) [3], as a linear approximation to the eigenfunctions of the nonlinear Laplacian Beltrami operator on the manifold, preserves the local manifold structures of the face image space. Face recognition is a

typical small sample size problem, and the number-of-dimensions to sample-size ratio is too high, so it is difficult for those subspace learning algorithms (PCA, LDA, LPP, *etc.*) to discover the intrinsic discriminative or geometrical structure. Even if the sample size were sufficient to estimate the intrinsic geometrical structure, coefficients of spatially smooth features (pixels) tend to be spatially rough. Deng, *et al.*, [4] proposed Spatially Smooth Subspace Learning (SSSL) model employing a Laplacian penalty to constrain the coefficients to be spatially smooth. In which, the face image is not considered as a  $w \times h$ -dimensional vector, but as a matrix or the discrete function defined on the  $w \times h$  lattice. The discrete Laplacian, which is a finite difference approximation to the second derivative operator, can be used as the basis functions to measure their smoothness along horizontal and vertical directions. Face image is classified in the lower dimensional subspace after it represented compactly with the SSSL model.

Wang, *et al.*, [5] constructed the complex face image by 8 bit-planes and recommended bit-plane generalized PCA (BGPCA) algorithm. The face's outline feature and the texture feature serve as the real and the imaginary part of the complex face image respectively. The outline feature is the weight sum of bit-planes 1, 2, 6-8, and the texture feature is weight sum of bit-planes 3-5. Dai [6] proposed that two-dimension PCA (2DPCA) is performed on the new real face images constructed with 8 bit-planes.

Wright [7] regarded face recognition as a globally sparse representation problem and defined the Randomfaces according to the sparse presentation theory in compressive sensing. Since then, sparse representation classification has gained wide-spread attention due to its prominently high recognition rate. Huang [8] proposed an improved method to make sparse representation invariant to image-plane transformation by simultaneously recovering the image plane transformation and sparse representation when the test image was not aligned with the model images. The method SRC [9] introduced an identity matrix  $I$  as an occlusion dictionary to code the outlier pixels. Yang [10] used Gabor features in SRC and proposed GSRC approach which greatly decreases the size of occlusion dictionary and improves the recognition accuracy. Yang [11] regarded the sparse coding as a sparsity-constrained robust regression problem, and the robust sparse coding (RSC) was proposed. Motivated by low-rank matrix recovery and completion, Ma [12] integrated rank minimization into sparse representation for dictionary learning and proposed discriminative low-rank dictionary learning for sparse representation (DLRD\_SR). The sparse noises are separated from the signals while simultaneously optimizes the dictionary atoms to reconstruct the denoised signals. In this algorithm, the over completed dictionary consists of the sub-dictionaries whose atoms are the samples from the same class. Since samples belongs to the same pattern are linearly correlated, the sub-dictionaries should be approximately low-rank. Minimizing the rank of the sub-dictionary separates the noises in the training images and makes the sub-dictionary compact. The over-completed dictionary is discriminative power because of the minimizing correlation between the sub-dictionary and the samples from other classes. Zhang [13] exploited local structure information among features in an image for the classification based on the image-level and recommended the low-rank sparse coding (LRSC) method which casts the feature coding problem as a low-rank matrix learning problem. Employing the similarities among local features lying in the same spatial neighborhood, the LRSC seeks an accurate joint representation of these local features *w.r.t.* a codebook that satisfies the locality property. The codes of local features are sparse and low-rank, which encourages that only a few (but the same) visual words are used to represent all features in a local region. As a robust sparse coding method, LRSC mines correlations among different local features and obtains the better coding results than learning each feature individually.

These mentioned above approaches based on the sparse representation are all time-consuming because of  $l_1$ -norm minimization, Zhang [14] proposed a collaborative representation based classification with regularized least square (CRC\_RLS) algorithm,

this novel face recognition algorithm prominently improves the recognition speed, but its recognition rate is inferior to RSC. Zhu [15] conducted CRC on patches and proposed the patch based CRC algorithm to solve the small sample size problem in face recognition.

Current researches have indicated that combining decisions of multiple classifiers can achieve better performance [16-18], in which voting strategy including majority voting and plurality voting is one of the most widely used combination strategies. Lin [19] presented the rationale behind plurality voting and demonstrated plurality voting is consisted with the Bayesian criterion and it outperforms majority voting in achieving a desirable trade-off between identification rate and rejection rate. Paul Watta [20] both theoretically and experimentally analyzed the performance of the plurality voting based ensemble classifier. Zhu [15] partitioned the query image into a set of overlapped patches and each patch is collaboratively represented over the corresponding set of patches of all training samples, and voting is used for the final decision making.

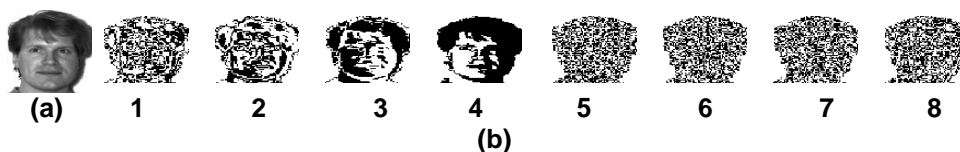
Inspired by the bit-plane images, collaboration and voting mechanism, we want to exploit the discriminative power of each bit-plane image, to find an approach to construct the image that contains more discrimination information than the original gray image, and to explore the application of plurality voting in face recognition jointing the bit-plane images and the collaboration representation classification. So this paper presents a novel face recognition approach using voting and collaborative representation theories based on binary bit-plane images. In this algorithm, all gray face images are firstly equalized and decomposed into 8 binary bit-planes, and then CRC\_RLS is performed on the five bit-planes which contain more discriminative information respectively and five identities of each test sample are obtained in this stage. Finally, voting mechanism is employed to decide the true identity of the test sample. Voting is failure when there are more than one identities get the largest number of votes for one test image. In this case, the virtual weight sum face image of the test image should be constructed by its 8 bit-plane images and CRC\_RLS is performed again on the virtual face database.

The rest of this paper is organized as follows: Section 2 explains image preprocessing and the bit-plane image extraction, followed by the details of the collaborative representation in Section 3; Section 4 describes the proposed approach; Section 5 conducts the experiments, and the conclusion is presented in Section 6.

## 2 Image Preprocessing

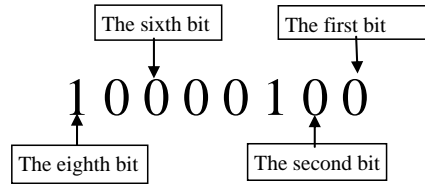
In this paper, image preprocessing includes histogram equalization and bit-plane extraction.

For the gray image with 256 gray levels, the values of the pixels can be expressed by 8 binary digits, so the image can be decomposed into 8 bit-planes by extracting each binary digit, from the least significant bit, bit-plane 1, to the most significant bit, bit-plane 8. Theory of image processing proves that the higher order bit-plane contains more visually significant data information, and the lower bit-plane images contain more texture information [21]. Figure 1 gives an example of one original gray face image and its 8 binary bit-plane images, it is clear that the higher-order bit-planes (1-4) carry the majority of the outline features; while the other bit-planes contain the more subtle details in the image.



**Figure 1. An Example of a Gray Face Image and its 8 Binary Bit-plane Images (a) Original Face Image (b) its Eight Bit-plane Images**

Here an example is given to explain how to build bit-plane images. Suppose  $I$  is a gray image with 256 gray levels, point  $P(x, y)$  is any pixel of  $I$ , the gray value of  $P$  is decimal integer 132, it can also be expressed by the binary string 10000100, the rightmost bit is called the first bit that is the lowest significant bit and the leftmost bit is called the eighth bit that is the highest significant bit. The bit order is explained in Figure 2.



**Figure 2. Explanation of the Bit Order**

Bit-plane image is obtained by extracting corresponding bit binary digit for all pixels of gray image  $I$ , for instance, the first bit-plane image is gotten by extracting the first binary digit (0) of gray value, and the eighth bit-plane image is gotten by extraction the eighth binary digit (1) of the gray value.

In BGPCA algorithm proposed by Wang [5], each bit-plane image is gray image whose gray value is associated with the order of the bit, but not binary image whose value is zero or one. Specific say, the pixels values of the first bit-plane image are  $0(0 = 0 \times 2^0)$  or alternatively  $1(1 = 1 \times 2^0)$ ; the pixels values of the second bit-plane image are  $0(0 = 0 \times 2^1)$  or alternatively  $2(2 = 1 \times 2^1)$ ; and the pixels values of the eighth bit-plane image are  $0(0 = 0 \times 2^7)$  or alternatively  $128(128 = 1 \times 2^7)$ . In a word, the gray values of the  $i^{\text{th}}$  bit-plane image are 0 or  $2^{i-1}$  alternatively.

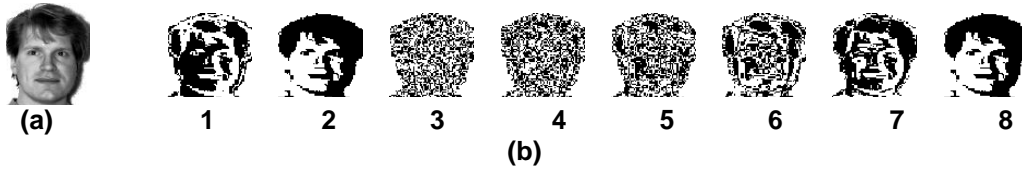
The complex face image is created by weighting 8 bit-plane gray images. Let  $V$  denotes the virtual image of  $I$ , which is given by  $V = V_r + jV_g = \sum_{m=0,1,5,6,7} \alpha_m B_m$ , where

$$V_r = \sum_{m=0,1,5,6,7} \alpha_m B_m, V_g = \sum_{m=2,3,4} \beta_m B_m, V_r \text{ and } V_g \text{ express the outline feature (namely the real part}$$

of  $V$ ) and the texture feature (namely the imaginary part of  $V$ ) respectively. Because the non-zero gray values ( $2^{i-1}$ ) of the higher bit-plane image are more larger than that of the lower bit-plane images, especially when the differences of the weight value  $\alpha_m$  or  $\beta_m$  are not bigger, so the higher order bit-plane image given the more contribution to the virtual image than the lower order bit-plane.

We believe this approach to construct the virtual image is not perfect and propose an improvement method to construct the real virtual image with the bit-plane images. In our proposed approach, the eight bit-plane images are all binary images, the pixels values of bit-plane images are 0 or 1 alternatively, the reason and advantages are discussed lately in Section 4.2 and in Section 6.

As a contrast enhancement technique, histogram equalization is able to enhance the global contrast and to make the subtle details clearer. In our experiments, all the face images are first equalized using the cumulative distribution function and then are decomposed into 8 bit-planes. The experiment results indicate that the information distribution in the bit-plane images is changed after equalization. Figure 3 gives the equalized image of the one shown in Figure 1 (a) and its 8 bit-planes. The outline features also emerge in the bit-planes 1 and 2, the top 3 bit-planes still contain abundance shape or outline information.



**Figure 3. Equalized Face Image and its Bit-plane Images (a) Equalized Face Image (b) its Eight Bit-plane Images**

### 3. Collaborative Representation for Classification

We assume to have  $K$  object classes and  $n_k$  training samples belonging to class  $k^{\text{th}}$ , so the total number of training samples is  $N = \sum_{k=1}^K n_k$ . Each face image is of the size  $w \times h$  and can be stacked as an  $m$ -dimensional vector ( $m = w \times h$ ). In the following sections, we denote the set made up by the training samples of the  $k^{\text{th}}$  class as the matrix  $\mathbf{X}_k = [\mathbf{x}_{k,1}, \dots, \mathbf{x}_{k,l}, \dots, \mathbf{x}_{k,n_k}]$ , where  $\mathbf{x}_{k,l}$  is an  $m$ -dimensional vector stretched by the  $l^{\text{th}}$  sample of the  $k^{\text{th}}$  class. The complete dictionary,  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K] \in \mathfrak{R}^{m \times N}$ , is the concatenation of the entire training samples of all  $K$  object classes. Let  $\mathbf{y}$  denote the test sample.

Sparse representation based classification (SRC) uses the compress sensing theory to find out the sparse representation of the test sample  $\mathbf{y}$  with  $\mathbf{X}$ . The algorithm emphasizes that the sparsity of the coefficient vector is very important to classify. Finally, solution of the coefficient vector is casted as the following  $l_1$ -minimization problem:

$$(\hat{\alpha}) = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{X}\alpha\|_2 < \varepsilon \quad (1)$$

$l_1$ -norm minimization is time-consuming and the computation speed is decreased sharply with the row-dimension or column-dimension of  $\mathbf{X}$  increasing. Most of the previous works focus on reducing the computational load via  $l_1$ -norm minimization, while the role of collaboration representation between classes is not paid enough attention to. Zhang [14] formally proposed the collaboration representation for the first time, where query sample is also linearly encoded by the entire training samples from all classes. The difference of the SRC is that it concentrates on the ‘collaboration’ instead of ‘sparsity’ while estimating the representation coefficients vector. In order to introduce a certain amount of ‘sparsity’ to the coefficient vector,  $l_2$ -norm is used as constrained condition. The coefficient vector is the solution of the following equation  $\hat{\rho}$  :

$$(\hat{\rho}) = \arg \min_{\rho} \{ \|\mathbf{y} - \mathbf{X}\rho\|_2^2 + \lambda \|\rho\|_2^2 \} \quad (2)$$

Utilizing regularized least square, the unique solution of Eq.(2) can be easily gotten as  $\hat{\rho} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$ , where  $\lambda$  is regularization parameter, term  $\lambda \|\rho\|_2^2$  ensuring the stability and sparsity of coefficient vector  $\hat{\rho}$ . Assume

$$\mathbf{P} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \quad (3)$$

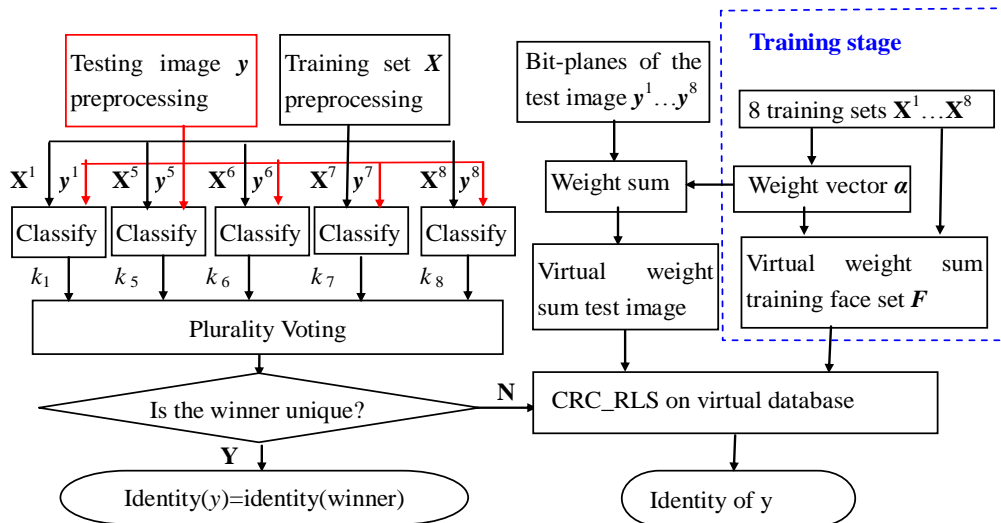
Then  $\mathbf{P}$  is independent of the test sample  $\mathbf{y}$ , so it can be regarded as the projection matrix obtained before testing. At the test stage, the query sample  $\mathbf{y}$  is only simply projected onto  $\mathbf{P}$  through  $\mathbf{P}\mathbf{y}$  at first, and then  $\mathbf{y}$  is classified with  $\hat{\rho}$ , evaluating which class leads to the minimum representation error. The procedures of the CRC\_RLS are summarized as follows.

**Algorithm CRC\_RLS:**

- a. Normalize the columns of the dictionary  $\tilde{X}$  to have unit  $l_2$ -norm, the normalized dictionary denoted by  $\tilde{X}$ , and then compute the project matrix  $P = (\tilde{X}^T \tilde{X} + \lambda I)^{-1} \tilde{X}^T$
- b. Normalize the query sample  $y$  and denote it as  $\tilde{y}$ , compute the coefficient vector  $\hat{\rho} = P \tilde{y} = (\rho_1, \dots, \rho_K)^T$ , where  $\rho_k \in \mathbb{R}^{L \times 1}$  is the coefficient vector associated with class  $k$ .
- c. Reconstruct  $K$  vectors  $\tilde{y}_k = \tilde{X}_k \rho_k$  ( $k=1 \dots K$ ).
- d. Compute the error  $r_k = \frac{\|\tilde{y} - \tilde{X}_k \rho_k\|_2}{\|\rho_k\|_2}$ . Theoretically, the larger elements of  $\hat{\rho}$  cluster round  $\rho_k$  if  $y$  is from class  $k$ , i.e.  $\|\tilde{y} - \tilde{X}_k \rho_k\|_2$  is smaller and  $\|\rho_k\|_2$  is larger, so the error term  $r_k$  possesses more discrimination information for classification.
- e. Identity of  $y$  is  $k = \arg \min_k \{r_k\}$ .

**4. Face Recognition using Plurality Voting and Collaboration Presentation**

In this paper, we present a new face recognition approach using plurality voting strategy and collaboration representation based on the bit-plane images (CRV\_BP). The block diagram of the proposed approach is shown in Figure 4.



**Figure 4. The Block Diagram of CRV\_BP**

This proposed method is comprised by a spare second identification and three indispensable parts, including preprocessing, off-line training and on-line recognition. The details are described in the following.

**4.1. Image Preprocessing**

All the images are preprocessed according to the operations described in Section 2. Eight training samples are obtained from the original training set, and each training set consists of the corresponding bit-plane images. Let  $x_{kl}^i$  denote the  $i^{\text{th}}$  bit-plane image from the  $l^{\text{th}}$  image of the  $k^{\text{th}}$  class ( $i=1,2 \dots 8$ ),  $\mathbf{X}^i = [x_{11}^i, \dots, x_{1n_1}^i, \dots, x_{kl}^i, \dots, x_{Kn_k}^i]$

denoting the  $i^{\text{th}}$  training matrix that contains all the  $i^{\text{th}}$  bit-plane images of the entire training samples.

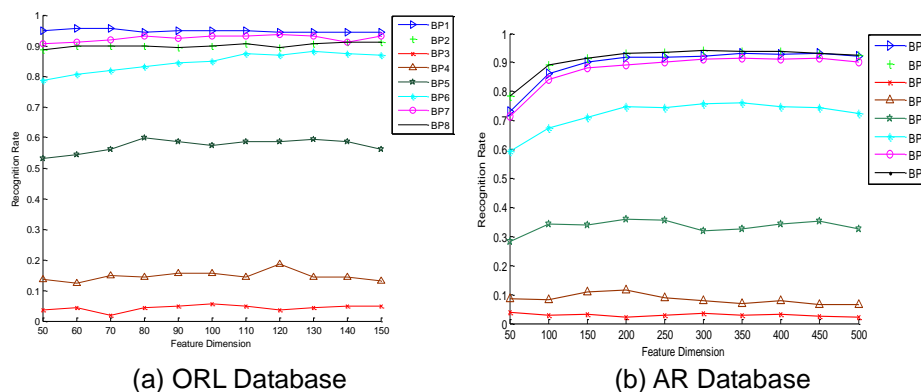
## 4.2. Previous Training

At this previous training stage, the weight vector must be computed and stored in case of need. The weight vector is affected by the right recognition rate that indicates the ability to discriminate. Thus, the first thing is to estimate the rate recognition rates of each bit-plane image.

**4.2.1. Estimation of the Recognition Rate Vector:** According to the assumption given in Section 3, there are  $K$  object classes and  $n_k$  training samples belonging to class  $k^{\text{th}}$ . Now we randomly select  $M$  ( $M < n_k$ ) images from the each subject to be as training samples, and the rest for testing. Let  $A^i = [A_1^i, A_2^i, \dots, A_K^i]$  be the training matrix corresponding to the  $i^{\text{th}}$  bit-plane, where  $A_k^i = [x_{k1}^i, \dots, x_{kM}^i]$  ( $k = 1, \dots, K$ ). The test sample set is denoted by  $S^i = \{x_{kt}^i \mid k = 1, \dots, K, t = M + 1, \dots, n_k\}$ . We replace  $X$  with  $A^i$  and compute the  $i^{\text{th}}$  project matrix  $P^i = [(A^i)^T A^i + \lambda I]^{-1} (A^i)^T$ . Let  $\xi_i$  indicate the right recognition rate corresponding to the  $i^{\text{th}}$  bit-plane, it is calculated through replacing  $y$  with each element  $x_{kt}^i$  of  $S^i$  and repeating the step (b)~(e) of the algorithm CRC\_RLS.

In order to boost the accuracy of the bit-plane recognition, the above motioned processing is repeated several times by transposing training images, certainly, the test images are also different in every experiment. The average of these recognition rates is as the end recognition rate  $\xi_i$ .

The recognition rate vector is gotten as  $\xi = (\xi_1, \dots, \xi_8)$  by repeating the above mentioned procedure, *i.e.*, the integer  $i$  varies from 1 to 8. The result of a certain experiment are presented in Figure 5, which shows the variation in each bit-plane recognition rate with the number of principal components retained, where the BP1, BP2 ... BP8 are abbreviations for the bit-plane 1, bit-plane 2...and bit-plane 8, respectively. In which bit-plane 2 and bit-plane 8 have the same recognition rate no matter how many features are retained, so the two curves, BP2 and BP8, overlap. The experimental results are similar even the training images are changed randomly, that is to say, this experimental results are general and they are less associated with the experimental data.



**Figure 5. Variation in Recognition Rate of Each Bit-plane with the number of Principal Components Retained (a) On the ORL Database, (b) On AR Database**

From the experimental results and the Figure 5, we can draw several conclusions: (1) The recognition rates of BP1, BP2, BP7 and BP8 are larger than the others bit-plane recognition rates. The recognition rates of BP3 and BP4 very lower, which are less than 0.2. The recognition rate of BP5 is also lower though it is greater than 0.3 in most cases. The recognition rate of BP6 is greater but it is never greater than 0.9. (2) Eight recognition rates tend to be stable when the feature dimension varies in a wide range. In Figure 5 (a), when the dimensionality of the data varies between 80 and 130, the variation of recognition rates is almost zero for BP1, BP2, BP7 and BP8; for the other four bit-planes, these changes are gently. The same things can be found in our all experiments, the difference lies in the data dimensionality. For example, for the AR data, the corresponding dimensionality interval becomes (150,500), which shown in Figure 5(b). (3) The proportion of eight recognition rates is basically stable though each individual recognition rate changes with the increasing of the feature dimensionality.

In our experiments, the dimensionality of the data increases from the start value to the end value according to a certain step, the start value, the end value and the step are related to the number of the training images. For instance, in the ORL database experiment, there are only 160 training images at this previous training stage, so we selected dimensionality of the data between 50 and 150, the growth step is 10. While in the AR database experiment, there are 500 training images, therefore dimensionality of the data varies from 50 to 500 with the growth step of 50. From Fig. 5, the right recognition rates are increasing with the dimension increased at first, but it tends to be stable soon, and then the recognition rates are stable in a wide range. So it is an easy job to determine the weight vector automatically by the variation rate of recognition rates. The recognition rate vector  $\xi = (\xi_1, \dots, \xi_8)$  is defined the average of its elements except for BP3 and BP4 reaches maximum.

**4.2.2. Estimation of Weight Vector:** Appropriate weight coefficient will be assigned to each bit-plane so that the constructed virtual image contains more discrimination information. We believe the bit-planes with the higher recognition rate carry more discrimination information for classification, while the bit-planes with very low recognition rate are unworthy to classify. So the weight coefficient should be related to the recognition rate. In addition, we have already known that the higher order bit-planes contain more discrimination information than the lower order bit-planes from Section 2. Taking these two factors into account, the weight coefficient is obtained by the Eq. (4), and weight vector being  $\alpha = (\alpha_1, \alpha_2 \dots, \alpha_8)$ .

$$\alpha_i = \begin{cases} 0 & \xi_i < 0.25 \\ i \times \xi_i^2 & \xi_i \geq 0.25 \end{cases} \quad (i = 1 \dots 8) \quad (4)$$

Define the virtual weight sum image of the original gray image  $x$  as

$$x_f = \sum_{i=1}^8 \alpha_i x^i \quad (5)$$

Here  $x^i$  is the  $i^{\text{th}}$  bit-plane image gotten from the image  $x$ ,  $x_f$  is the virtual weight sum image. It is worth mentioning that the relative amplitude of each element in  $\alpha$  to others is critical to the weight virtual image, while the absolute value of the elements is trivial.

According to the Section 2, each bit-plane images  $x^i$  is binary image and they present their self different contribution to  $x_f$  but the differences is limited, hence the virtual weight sum image included more discrimination information than the original image or any of binary bit-plane images.



**4.2.3 Construction of the Virtual Dictionary:** The completed weight sum dictionary  $F$  and virtual project matrix  $V$  can be computed as  $F = \sum_{i=1}^8 \alpha_i X^i$  and  $V = (F^T F + \lambda I)^{-1} F^T$ , respectively.

### 4.3 Real-time Recognition

**4.3.1 Pre-processing of Testing Image:** At first, we perform histogram equalization and bit-plane extraction on the query image  $y$  and obtain its eight binary bit-plane images  $y^i (i = 1, 2, \dots, 8)$ . From the image processing theory described in Section 2, we know that the 3rd and the 4th bit-planes contain more texture information which is unhelpful for classification, and that the experiment results from the off-line training also verified the recognition rates corresponding to the two bit-planes are always less than 0.25. Therefore, the 3rd and the 4th bit-planes are discarded in the procedure of plurality voting that will be described later. The training results also indicated that bit-plane 2 and bit-plane 8 have the same recognition rate whenever and they will bring the same identity for the test image, hence the 2nd bit-plane is also discarded when voting for saving recognition time. Left five bit-planes vote for the identity of test image according to the following methods.

**4.3.2 Identity Decision of  $y^i$ :** In CRC\_RLS, replace  $X$  and  $y$  with  $X_i$  and  $y^i (i = 1, 5, 6, 7, 8)$  respectively, and the identity of  $y^i$  is computed,  $k_i$ , an integer between 1 to  $K$ , denoting the identity of  $y^i$ .

**4.3.3 Plurality Voting:** For most problems with several solutions, voting can be used to improve the system's reliability or accuracy. Generally, the voting strategies include the majority voting and the plurality voting. The plurality voting is to select the candidate which receives the highest number of votes to be the winner, in which each classifier computes its output and casts a vote for that class, a decision network counting up the votes and choosing the class with the most votes. Now, we use the 5 recognition results and the plurality voting strategy to decide the identity of test sample  $y$ .

Assume there are 5 vectors whose length equals to  $K$  (number of the classes)  $v_1, v_2, \dots, v_5 \in \mathbb{R}^K$  and let  $d_1 = k_1, d_2 = k_5, d_3 = k_6, d_4 = k_7, d_5 = k_8$ . The values of  $v_p (p = 1, 2, \dots, 5)$  are assigned by Eq. (6):

$$v_p(q) = \begin{cases} 1 & d_p = q \\ 0 & \text{others} \end{cases} \quad (6)$$

The total number of votes received by class  $m$  is presented by

$$v_q = \sum_{p=1}^5 v_p(q) \quad (7)$$

This number of votes  $v_q$  presents measure similarity between the input and class  $q$ . The test sample and the individual who received the most votes, *i.e.*, the winner, have the same identity.

Let vector  $v = (v_1, v_2, \dots, v_K)$ , and suppose  $v_Q$  is the only largest element, *i.e.*  $v_Q = \max\{v_q, q = 1, 2, \dots, K\}$ . As the subscript of  $v_M$ ,  $Q$  denotes the identity of the test image. If there are more than one subjects having the max number votes, for example,  $v_{Q_1} = v_{Q_2} > v_q (q = 1, 2, \dots, K, q \neq Q_1, q \neq Q_2)$ , we believe the plurality voting fails. In this case, it is necessary to classify the test image  $y$  according to the second identification.

#### 4.4. The Second Identification

When the plurality voting fails, the second recognition will be used to determine the true identity of the test image  $y$ . The first thing is to create the virtual test image employing the weight vector  $\alpha$  and 8 bit-plane images of  $y$ . The virtual test image from  $y$  is defined as  $y_f = \sum_{i=1}^8 \alpha_i y^i$ . Let  $F$  and  $y_f$  take place of  $X$  and  $y$  respectively, the true identity of  $y$  will be decided by CRC\_RLS algorithm at last.

### 5. Experimental Results

In this section, we have tested the performance of the proposed algorithm (CRV\_BP) and compared it with PCA, RSC CRC\_RLS on the AR database and ORL database. In all experiments, the regularization parameter  $\lambda$  is set as 0.001 empirically, and the classifier is near subspace classifier. In PCA, assign the test image  $y$  to subject  $k$  if the distance from  $y$  to the subspace spanned by all images  $\mathbf{x}_k = [x_{k,1}, \dots, x_{k,l}, \dots, x_{k,m_k}]$  of the subject  $k$ ,  $r_k = \min_k \|\tilde{y} - \text{avg}(\tilde{X}_k)\|_2, (k = 1, \dots, K)$ , is the smallest among all subjects.

Here  $\tilde{y}$  is the low-dimension feature image of  $y$ ,  $\text{avg}(\tilde{X}_k)$  is the average feature image of the  $k^{\text{th}}$  subject.

#### 5.1. ORL Face Database

ORL database contains 400 grayscale images from 40 individuals, each individual having 10 images. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/ no glasses). The size of the images is  $112 \times 92$ . Experiments demonstrate image resolution has less influence on the recognition rate, so all the face images are normalized to  $64 \times 64$  pixels for saving recognition time in our experiment. For each individual, 6 images are used for training, the rest 4 images are used for testing, total numbers of training samples and testing samples are 240 and 160 respectively. This experiment is repeated several times through choosing the different training samples, and the testing samples are also different certainly. Experiment results shown that the sample selection is little effect on the experimental results.

At the previous training stage, we randomly select 4 from the 6 training images of each subject for training, and the rest training images for testing. Figure 5 (a) shows the variation of each bit-plane recognition rate relative to the dimension of the features. As discussed in Section 4.2, this procedure is also repeated several times and it is not difficult to find out an appropriate recognition rate vector. With selected recognition rate vector as  $\xi = (0.9500, 0.9063, 0.0500, 0.1438, 0.5875, 0.8750, 0.9313, 0.9063)$ , weight vector is computed by Eq. (4) as  $\alpha = (0.9025, 1.6426, 0.017258, 4.5938, 6.0706, 6.5703)$ . Table 1 gives the experiment results.

It is necessary to construct the virtual test image using  $\alpha$  and perform the CRC\_RLS on the virtual database when the plurality voting fails. For 160 testing images, there are 2 images could not be recognized by plurality voting described in Section 4.3 when the feature dimension equals to 95. In this case, the spare second identification is employed to determine the pattern of test image.

**Table 1. The Recognition Rates on ORL Database (%)**

Feature dimension	85	90	95	100	150	400	1000
PCA	86.88	87.5	88.12	87.5	87.5	--	--
SRC	85	79.37	77.5	83.13	88.75	92.5	86.88
CRC_RLS	91.88	92.50	93.13	92.50	90.63	--	--
CRV_BP	<b>94.37</b>	95.63	<b>96.88</b>	95.63	95.63	--	--

### 5.2. AR Face Database

AR database contains 3288 color face images of 116 individuals, including frontal views of faces with different facial expressions, lighting conditions and occlusions. In this paper, we utilize the light component to be the gray images and crop the face portion of the images whose size are  $60 \times 43$ .

Taking 1400 images from 100 persons (50 men and 50 women) to experiment, each person has 14 images. For each individual, 7 images are used for training, the rest 7 images are used for testing, and the total number of training and testing sample are both 700. This experiment is repeated several times through choosing the different training samples, and the testing samples are also different certainly. Experiment results shown that the sample selection is little effect on the experimental results.

At the previous training stage, for each subject, we randomly choose 5 from the 7 training images as training samples, and the rest 2 training images for testing. Figure 5 (b) shows the variation of each bit-plane recognition rate relative to the dimension of the features. With selected recognition rate vector as  $\xi = (0.9299, 0.9313, 0.0258, 0.0644, 0.3534, 0.7454, 0.9142, 0.9313)$ , weight vector is computed by Eq. (4) as  $\alpha = (0.8647, 1.7346, 0, 0, 0.6450, 3.3337, 5.8503, 6.9386)$ . The recognition rates obtained from the 700 testing images are shown as Table 2.

**Table 2. The Recognition Rates on AR Database (%)**

Feature dimension	54	120	300
PCA	68.00	70.10	71.24
SRC	83.30	89.5	93.28
CRC_RLS	80.50	90.00	93.85
CRV_BP	83.83	<b>92.85</b>	<b>96.14</b>

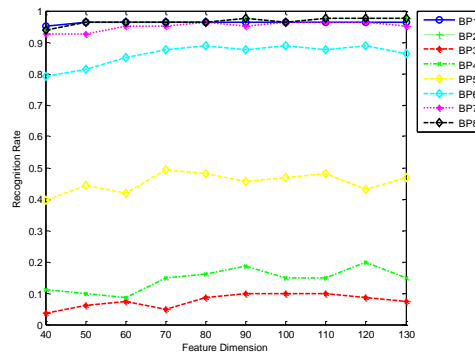
When the feature dimension is 300, 21 testing samples of 700 testing images could not be recognized by plurality voting and their identities are decided by the spare second identification.

### 5.3. FERET Face Database

For FERET database, we selected 216 grayscale images from 27 individuals, each individual having 8 frontal face images with different facial expressions and lighting conditions. Sizes of the images are  $40 \times 35$ . For each individual, 6 images are used for training, the rest 2 images are used for testing, and the total number of training and testing sample are 162 and 54 respectively. This experiment is repeated several times through choosing the different training samples, and the testing samples are also different certainly. Experiment results shown that the sample selection is little effect on the experimental results.

At the previous training stage, for each subject, we randomly choose 5 from the 6 training images as training samples, and the rest one training image for testing. Figure 6

shows the variation of each bit-plane recognition rate relative to the dimension of the features. With selected recognition rate vector as  $\xi = (0.9630, 0.9753, 0.0864, 0.1975, 0.4321, 0.8889, 0.9630, 0.9753)$ , weight vector is computed by Eq.(4) as  $\alpha = (0.9273, 1.9025, 0, 0, 0.9335, 4.7407, 6.4911, 7.6098)$ .



**Figure 6. Variation in Recognition Rate of Each Bit-plane with the Number of Principal Components Retained on the FERET Database**

When the feature dimension is 120, there is no testing samples could not be recognized by plurality voting. The recognition rates obtained from the 54 testing images are shown as Table 3.

**Table 3. The Recognition Rates on FERET Database (%)**

PCA	88.89
SRC	96.30
CRC_RLS	94.44
CRV_BP	98.15

#### 5.4 Comparison of Recognition Speed

At the end of this paper, we compare the recognition speed of the proposed method CRV\_BP with the competing approaches SRC and CRC\_RLS. The results of ORL and AR are listed in Table 3 and Table 4 respectively. In the below Tables, the optimal feature-dimension, which may be different for each approach, is the minimum feature-dimension when the recognition rate reaches highest. As a parameter to evaluate the recognition speed, the testing time is the average running time to accomplish one testing under the optimal feature-dimension.

**Table 4. Speed on ORL Database**

Method	Recognition rate (%)	Optimal feature-dimension	testing time (s)
SRC	92.5	400	0.3452
CRC_RLS	93.13	95	0.0087
CRV_BP	96.88	95	0.0325

**Table 5. Speed on AR Database**

Method	Recognition rate (%)	Optimal feature -dimension	Testing time (s)
SRC	93.30	300	1.8258
CRC_RLS	93.85	300	0.0051
CRV_BP	96.14	300	0.0301

From the above experimental results, we can draw the following conclusions. First, SRC is time-consuming, its testing time are more than 10 times of the others. The reasons are that SRC performs dimension reduction with the random projection matrix and that SRC seeks the sparse coefficient vector by Eq. (1) that is very complex and time-consuming. It is infeasible to apply SRC to real-time applications.

Second, CRC\_RLS approach simply computes the coefficient vector of testing image  $y$  through projecting  $y$  onto  $P$  via  $P \times y$ , and the projection matrix  $P$  can be obtained in the training stage before the testing. It makes CRC\_RLS very fast.

Finally, the proposed method CRV\_BP consists of image histogram equalization, bit-plane extraction, running CRC\_RLS several times and plurality voting, so it may involve constructing virtual weight sum test image and performing the procedure CRC\_RLS again. Hence CRV\_BP spends more time than CRC\_RLS in recognizing. But its speed is faster than the SRC prominently, because CRV\_BP avoids the time-consuming Eq. (1) and the components of CRV\_BP are simple and time saving.

## 6. Conclusion

A new face recognition approach with the collaboration representation and plurality voting based on the bit-plane images has been discussed in this article. In this proposed approach, all original gray face images are equalized by cumulative distribution function, and five binary bit-plane images with more discrimination information vote to decide the identity of testing image. When the plurality voting fails, CRC\_RLS is performed again on the virtual weight sum face database to determine the true identity of test sample. The weight vector is related to the right recognition rate and the order of each bit-plane, and eight bit-plane images are all binary, *i.e.*, their pixel values are 0 or 1 alternatively, hence each bit-plane presents different contribution and the differences among these contributions are limited. These factors make the virtual weight sum image include more discrimination information than the original image or any of the binary bit-plane images. The bit-planes with less recognition information are discarded in the procedures of plurality voting and creating virtual image for sake of interference.

In the voting procedure, the third and the fourth bit-plane images are absent from voting because of their low right recognition rate; For the second and the eighth bit-plane images, one of them is selected to vote since they have the same recognition rate, and we select the eighth bit-plane image to vote. The plurality voting is successful in most cases, but if the plurality voting fails, the virtual face image of the testing image is constructed and as a new voter gives its critical decision for the true identity of the testing image. All these reasons mentioned above are helpful to increase the recognition accuracy and reduces the judgment time.

The test images are represented collaboratively by all the training images from all classes and classified individually under the  $l_2$ -norm constraint, which solves the “lack of samples” problem in face recognition and avoids the expensive  $l_1$ -norm minimization at the same time. The speed of proposed approach is prominently higher than the SRC.

The algorithm of CRV\_BP is promising, but if the expression or/and the pose changes intensely, the algorithm needs to be improved. In later works, we will pay attention to dealing with face images with intense variation in expression or/and pose.

## Reference

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition", *J Cognitive Neuroscience*, vol. 3, no. 1, (1991), pp. 71–86.
- [2] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", *IEEE T Pattern Anal Mach Intell*, vol. 19, no. 7, (1997), pp. 711–720.
- [3] X. He, S. Yan, Y. Hu and H. Zhang, "Learning a locality preserving subspace for visual recognition", In *Proceedings of IEEE International Conference on Computer Vision, Nice, France, (2003)*, pp. 385-392.
- [4] C. Deng, X. F. He, Y. X. Hu, J. W. Han and H. Thomas, "Learning a Spatially Smooth Subspace for Face Recognition", In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, (2007)*, pp. 1-7.
- [5] H. Y. Wang, Y. Leng, Z. Wang and X. Wu, "Application of image correction and bit-plane fusion in generalized PCA based face recognition", *Pattern Recognition Lett*, vol. 28, no. 16, (2007), pp. 2352-2358.
- [6] Y. Dai, "Recognition of humans in video sequence", *Dissertation, Southwest University, (2010)*.
- [7] A. Yang, J. Wright, Y. Ma and S. Shankar, "Feature selection in face recognition: A sparse representation perspective", *Technical Report UCB/EECS-2007-99, UC Berkeley, (2007) August 14*.
- [8] J. Z. Huang, X. L. Huang and D. Metaxas, "Simultaneous image transformation and sparse representation recovery", In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Anchorage, AK, (2008)*, pp. 1-8.
- [9] J. Wright, A. Yang, A. Ganesh, S. Sastry and Y. Ma, "Robust face recognition via sparse representation", *IEEE T Pattern Anal Mach Intell*, vol. 31, no. 2, (2009), pp. 210–227.
- [10] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary", In *Proceedings of ECCV'10 Proceedings of the 11<sup>th</sup> European conference on Computer vision: Part VI. Heraklion Crete Greece: Springer-Verlag Berlin, Heidelberg, (2010)*, pp 448-462.
- [11] M. Yang, L. Zhang, J. Yang and D. Zhang, "Robust sparse coding for face recognition", In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Providence, RI, (2011)*, pp. 625-632.
- [12] L. Ma, C. H. Wang, B. H. H. Xiao and W. Zhou, "Sparse Representation for Face Recognition based on Discriminative Low-Rank Dictionary Learning", In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Providence, RI, (2012)*, pp. 2586-2593.
- [13] T. Z. Zhang, B. Ghanem, S. Liu C. S. Xu and N. Ahuja, "Low-Rank Sparse Coding for Image Classification", In *Proceedings of IEEE International Conference on Computer Vision. Sydney, NSW, (2013)*, pp. 281-288.
- [14] L. Zhang, M. Yang and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition", In *Proceedings of IEEE International Conference on Computer Vision, Barcelona, (2011)*, pp. 471-478.
- [15] M. Zhu, L. Zhang, Q. Hu and S. Shiu, "Multi-scale Patch Based Collaborative Representation for Face Recognition with Margin Distribution Optimization", In *Proceedings of ECCV'12 Proceedings of the 12<sup>th</sup> European conference on Computer vision: Part I. Florence Italy: Springer-Verlag Berlin, Heidelberg, (2012)*, pp. 822-835.
- [16] M. Erp, L. Vuurpijl and L. Schomaker, "An overview and comparison of voting methods for pattern recognition", In *Proceedings of the eighth international workshop on frontiers in handwriting recognition, Canada, (2002)*, pp. 195–200.
- [17] T. K. Ho, J. J. Hull and S. N. Srihari, "Decision combination in multiple classifier systems", *IEEE T Pattern Anal Mach Intell*, vol. 16, no. 1, (1994), pp. 66–75.
- [18] J. Kittler, M. Hatef, R. Duin and J. Matas, "On combining classifiers", *IEEE T Pattern Anal Mach Intell*, vol. 20, no. 3, (1998), pp. 226–239.
- [19] X. Lin, S. Yacoub, J. Burns and S. Simske, "Performance analysis of pattern classifier combination by plurality voting", *Pattern Recognition Lett*, vol. 24, no. 12, (2003), pp. 1959–1969.
- [20] X. Mu, P. Watta and M. H. Hassoun, "Analysis of a plurality voting-based combination of classifiers", *IEEE International Joint Conference on Neural Networks IJCNN, (2008)*, pp. 304–309.
- [21] C. G. Rafael and E. W. Richard, "Digital Image Processing", *Prentice Hall, NJ, USA, (2003)*.

## Author



**Dongmei Wei**, was born in October 1978. Now she majors in pattern recognition and image segmentation.

