Neural Process Lett (2012) 35:233–244
DOI 10.1007/s11063-012-9214-4

Heteroscedastic Sparse Representation Based Classification for Face Recognition

Hao Zheng · Jianchun Xie · Zhong Jin

Published online: 5 February 2012

Abstract Sparse representation based classification (SRC) have received a great deal of attention in recent years. The main idea of SRC is to represent a given test sample as a sparse linear combination of all training samples, then classifies the test sample by evaluating which class leads to the minimum residual. Although SRC has achieved good performance, especially in dealing with face occlusion and corruption, it must need a big occlusion dictionary which makes computation very expensive. In this paper, a novel method, called heteroscedastic sparse representation based classification (HSRC), is proposed to address this problem. In the presence of noises, the SRC model exists heteroscedasticity, which makes residual estimation inefficient. Therefore, heteroscedastic correction must be carried out for homoscedasticity by weighting various residuals with heteroscedastic estimation. As for heteroscedasticity, this paper establishes generalized Gaussian model through which to estimate. The proposed HSRC method is applied to face recognition (on the AR and Extended Yale B face databases). The experimental results show that HSRC has significantly less complexity than SRC, while it is more robust.

Keywords Heteroscedastic · Face recognition · Sparse representation based classification (SRC) · Generalized Gaussian model (GGM)

1 Introduction

In recent years, there is an increasing trend of using face recognition technology, and many face recog-nition techniques have been developed. For unsupervised methods, e.g., princi-

H. Zheng (✉) · J. Xie · Z. Jin
School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China
e-mail: zhh710@163.com

H. Zheng
School of Mathematics and Information Technology, Nanjing Xiao Zhuang University, Nanjing 211171, People's Republic of China

pal component analysis (PCA) [13] and locality preserving projection (LPP) [11], the low-dimensional representa-tion should discover the structure information of the data point cloud [4]. For supervised classifica-tion problem, e.g., linear discriminant analysis (LDA) [12,15] and maximum margin crite-rion (MMC) [16], the reduced low-dimensional features should contain most discriminative informa-tion based on the labeled data. But their performances for noise images are not ideal.

In recent years, sparse representation has aroused broad interest due to its great success in image processing [6,18], and it has also been used for face recognition(FR). The basic idea of sparse representation is to represent a given test image using a small number of atoms parsimoniously chosen out of an over-complete dictionary. The sparsity of the coding coefficient can be measured by l_0-norm, which counts the number of nonzero entries in a vector. However the combinatorial l_0-norm minimization is an NP-hard problem, so the approximation methods need to be adopted. Several methods have been proposed recently, e.g., match pursuit(MP) and basic pursuit(BP).

Based on MP method [10,19,28], application-based attempts were drawn to look at special features such as nose localization in face recognition process. In this face recognition application, Phillips [14] generalized Mallat and Zhang's matching pursuit algorithm to simultaneously decompose multiple images for pattern recognition application and introduced matching pursuit filters (MPFs) as an adapted wavelet expansion build from a training set of the object. Francois [7] also proposed a new shape descriptor on the basis of the use of matching pursuit as a shape analysis tool, and extended it by introducing a scale-space approach and an affine invariant dictionary. In 2010, for overcoming the large size of the training matrix limitations, Patel [24] proposed an automatic target recognition algorithm based on learning class supervised dictionaries for simultaneous sparse signal representation and classification. But these methods are not ideal for recognition of images with occlusion and corruption.

Based on BP method [2], Sparse representation based classification (SRC) was reported by Wright et al. [30] for robust face recognition in 2009. In Wright et al.'s pioneer work, the training face images are used as the dictionary of representative samples, and an input test image is coded as a sparse linear combination of these sample images via l_1-norm minimization. The results in [5,30,32] of SRC were exciting in FR, which could lead to high classification accuracy, especially well handling the problems of face occlusion and corruption. Furthermore, Yang and Zhang [31] used Gabor features in SRC to reduce computation complexity and improve recognition rate with a learned Gabor occlusion dictionary. However, the main shortcomings of these methods is that a large number of training images are required for good recognition performance. Especially when the number of atoms required in the orthogonal occlusion matrix is very large, we get a very big occlusion dictionary, which makes the sparse process very computationally expensive and the real-time processing very difficult.

In order to avoid the problem of the occlusion dictionary, a novel method called heteroscedastic sparse representation based classification (HSRC) is proposed. In real world, such noises as occlusion and corruption make SRC model exist heteroscedasticity which makes residual estimation inefficient. Therefore, Heteroscedastic correction must be carried out for homoscedasticity by weighting various residuals with heteroscedastic estimation. Because the residual distribution can be approximated by generalized Gaussian model (GGM), we can obtain heteroscedasticity through maximum likelihood (ML) estimation of residuals. Different from the approximate decomposition of MP methods and the l_1-regularized minimization of BP methods, the proposed HSRC method adopts l_2-regularized least square, introduce residuals weights which are computed approximating these residuals by using a zero-mean generalized Gaussian density. The proposed HSRC method was extensively evaluated on FR

in different conditions, including variations of illumination, expression, occlusion, corruption. The experimental results demonstrated that HSRC outperforms significantly previous state-of-the-art methods. In particular, HSRC could achieve very high recognition rate but with low computational cost.

The rest of this paper is organized as follows: Sect. 2 briefly reviews SRC. Then proposed HSRC algorithm is presented in Sect. 3. Section 4 conducts experiments to validate the proposed method and Sect. 5 concludes the paper.

2 Sparse Representation Based Classification [30]

Given a signal (or an image) $y \in \Re^m$, and a matrix $A = [a_1 \ldots, a_2, a_n] \in \Re^{m \times n}$ containing the elements of an overcomplete dictionary [23] in its columns, the goal of sparse representation is to represent y using as few entries of A as possible. This can be formally expressed as follows:

$$\hat{x} = \arg\min_x ||x||_0 \, s.t. \, y = Ax \tag{1}$$

where $x \in \Re^n$ is the coefficient vector, and $||x||_0$ is the l_0-norm which is equal to the number of non-zero components in x. However, this criterion is not convex, and finding the sparsest solution of Eq. 1 is NP-hard. Fortunately this difficulty can be by passed by convexizing the problem and solving

$$\hat{x} = \arg\min_x ||x||_1 \, s.t. \, y = Ax \tag{2}$$

where l_1 is used instead of l_0. It can be shown that if the solution x sought is sparse enough, the solution of l_0 minimization problem is equal to the solution of l_1 minimization problem. This problem is convex and can be solved with variety of methods, e.g., Basis Pursuit [3] method which uses an interior point method to find a solution.

Finally, for each class j, let $\delta_j : \Re^n \to \Re^n$ be the characteristic function which selects the coefficients associated with the j-th class. Using only the coefficients associated with the j-th class, one can approximately reconstruct the test sample y as $\hat{y} = A\delta_j(\hat{x})$, then classify y based on these approximations by assigning it to the class that minimizes the residual:

$$r_j(y) = ||y - A\delta_j(\hat{x})||_2, \, for j = 1, \ldots, k. \tag{3}$$

If $r_l(y) = \min r_j(y)$, y is assigned to class l.
Now suppose that the face image is partially occluded or corrupted, the problem can be expressed as follows:

$$\hat{x} = \arg\min_x ||x||_1 \, s.t. \, y = Ax + \varepsilon \tag{4}$$

where ε is residual. We can approximately reconstruct the test sample y as $\hat{y} = A\delta_j(\hat{x}) + \hat{\varepsilon}$, then compute the residuals:

$$r_j(y) = ||y - A\delta_j(\hat{x}) - \hat{\varepsilon}||_2, \, for j = 1, \ldots, k. \tag{5}$$

If $r_l(y) = \min r_j(y)$, y is assigned to class l.

3 Heteroscedastic Sparse Representation Based Classification (HSRC)

In this section, we show how heteroscedastic sparse representation based classification can be used for FR, and discuss why the proposed method is much more effective and efficient than state-of-art FR methods.

3.1 Heteroscedastic Correction for Homoscedasticity

Equation 4 can also be expressed as the original model

$$y = Ax + \varepsilon \qquad \text{s.t. } \|x\|_1 \leq \rho(\rho \text{ is a constant}) \tag{6}$$

where $\varepsilon = (\varepsilon_1, \varepsilon_{2.}, \ldots, \varepsilon_i, \ldots, \varepsilon_n)$, $\text{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0 i \neq j \\ 1 i = j \end{cases}$ $i, j = 1, 2, \ldots, n$.

It can be seen that Eq. 6 is equivalent with objective function as follow:

$$\hat{x} = \arg\min_x \left\{ ||y - Ax||_2^2 + \lambda \|x\|_1^2 \right\}$$

If $E(\varepsilon_i) = 0 \quad Var(\varepsilon_i) = \sigma^2 i = 1, 2, \ldots, n$.
Then covariance matrix of the ε

$$E(\varepsilon\varepsilon^T) = \sigma^2 I = \sigma^2 \begin{bmatrix} 1 & & 0 \\ & 1 & \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

Thus the original mode exists homoscedasticity which means that the residuals in the general model are identically distributed, i.e. that they have the same variance.
But in practice there are usually different variances due to noises, i.e.

$$Var(\varepsilon_i) = \sigma_i^2 i = 1, 2, \ldots, n$$

then covariance matrix of the ε

$$E(\varepsilon\varepsilon^T) = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix} \neq \sigma^2 I$$

we consider the original mode exists heteroscedasticity which make estimator of coefficients no longer efficient. Thus one possible way is to correct the heteroscedasticity through weighted least square which make residuals of large variance have low weight values and residuals of small variance have high weight values. Heteroscedasticity needs to be transformed to homoscedasticity [9,17,22].

We assume $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$
Let $W = diag(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_n^{-1})$
Now the following transformed model can be obtained

$$y^* = A^* x + \varepsilon^* \tag{7}$$

where $y^* = Wy$, $A^* = WA$, $\varepsilon^* = W\varepsilon$
Consider $E(\varepsilon^*(\varepsilon^*)^T) = E(W\varepsilon\varepsilon^T W^T) = WE(\varepsilon\varepsilon^T)W^T = W\Sigma W^T = I$
we can see that the transformed model is homoscedastic.

3.2 Coefficients of the Transformed Model

Hence in the new model the residual sum of squares is:

$$J = \sum_{i=1}^{n} (\varepsilon_i{}^*)^2 = \sum_{i=1}^{n} (\frac{\varepsilon_i}{\sigma_i})^2 \tag{8}$$

The estimation of sparse coefficients can be obtained by minimizing the J:

$$\begin{aligned} \hat{x} &= \arg\min_x J = \arg\min_x \sum_{i=1}^{n} (\frac{1}{\sigma_i}\varepsilon_i)^2 = \arg\min_x \sum_{i=1}^{n} (\frac{1}{\sigma_i}(y_i - a_i x_i))^2 \\ &= \arg\min_x (y - Ax)^T W (y - Ax) \end{aligned} \tag{9}$$

In order to keep the coefficients sparsity, the sparsity term $||x||_1$ must be added. However the l_1- minimization needs much computational cost due to the sparsity of l_1- norm. We consider there might be other minimization method which can substitute for l_1- minimization method. Recent research works [25,27,33] have questioned the use of sparsity for image classification. Shi et al. [27] showed that the sparsity assumption which underpins much of this work is not supported by the data, and a simple l_2- norm to the face recognition problem is not only significantly more accurate than the state-of-the-art approach, it is also more robust, and much faster. Moreover Zhang et al. [33] have shown that it is not necessary to impose the l_1- sparsity constraint, while the l_2- norm regularization performs equally well. In order to increase the efficiency of the proposed method, we substitute l_1- minimization with l_2- minimization as follow:

$$\hat{x} = \arg\min_x \left\{ (y - Ax)^T W (y - Ax) + \lambda \left\| x \right\|_2^2 \right\} \tag{10}$$

where λ is positive scalar.

We assume $Q = (y - A\hat{x})^T W (y - A\hat{x}) + \lambda \left\| \hat{x} \right\|_2^2$

$$\frac{\partial Q}{\partial \hat{x}} = -2A^T W (y - A\hat{x}) + 2\lambda\hat{x} = 0$$

Thus the coefficient $\hat{x} = (A^T W A + \lambda)^{-1} A^T W y$

3.3 Heteroscedastic Weights

How do we obtain W? First, we can construct new model to obtain σ_i. Due to occlusions, corruptions and expressions variations occur in the face images, residual ε doesn't follows Gaussian or Laplacian distribution. Fortunately a good probability density functions (PDF) approximation for the residual may be achieved by adopting a zero-mean generalized Gaussian density [1,21,26], which is defined as

$$p(\varepsilon_i; \sigma_i^2, \beta) = \frac{\beta\eta(\sigma, \beta)}{2\Gamma(1/\beta)} \exp\{-[\eta(\sigma, \beta)|\varepsilon_i|]^\beta\} \tag{11}$$

Where σ_i and β denote the variance and the shape parameter of the distribution respectively. Γ is the gamma function given by $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt . x > 0$. In addition, $\eta(\sigma, \beta)$is given by

$$\eta(\sigma, \beta) = \frac{1}{\sigma_i}\left[\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}\right]^{1/2}.$$

From above we can see that the GGM contains the Gaussian and Laplacian model as special cases, using β=2 and β=1, respectively.

Furthermore, in [29] evaluation of accuracy of estimating for GGM models among classic statistical methods shows that the ML estimator is significantly superior. We now describe an ML estimator for GGD.

We define the likelihood function of the residual $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ having independent component as $L(\varepsilon; \sigma_i, \beta) = \ln \prod_{i=1}^{n} p(\varepsilon_i; \sigma_i, \beta)$.

The following likelihood equations have a unique root in probability, which is indeed the maximum likelihood estimator

$$\frac{\partial L(\varepsilon; \sigma_i, \beta)}{\partial \sigma_i} = \frac{1}{\eta}\frac{\partial \eta}{\partial \sigma_i} - \beta(\eta\varepsilon_i)^{\beta-1}|\varepsilon_i|\frac{\partial \eta}{\partial \sigma_i} = 0 \tag{12}$$

The ML method given the estimates of parameters σ_i as follows:

$$\sigma_i = \left[\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}\right]^{1/2} \beta^{1/\beta}|\varepsilon_i| \tag{13}$$

Because experiments show the values of β is set to [0.2, 2], a reasonable β can be found to make good recognition performance. Then W can be obtained by σ_i. It is clear that the values of weight are inversely proportional to the values of the residual. For simplicity, initial ε_i is obtained by ordinary least squares estimation.

3.4 Heteroscedastic Sparse Representation Based Classification Algorithm (HSRC)

Since initial residual is an estimator which is not optimal, the implementation of HSRC is an iterative process. Considering the fact that the residual sum of squares ε^2 decreases in each iteration, and it is non-negative value, the proposed HSRC algorithm will converge. When the difference of ε^2 is small enough, the convergence is stopped. It can be formulated as follow:

$$\left\|\varepsilon^{(t)}\right\|_2^2 - \left\|\varepsilon^{(t+1)}\right\|_2^2 / \left\|\varepsilon^{(t)}\right\|_2^2 < \eta$$

where η is a small positive scalar, t is the t-th of the iteration. In order to verify the convergence of the HSRC algorithm, Experiments on FR database were done. It is straightforward that the residual sum of squares ε^2 decreases as the number of iterations increases, as illustrated in Fig. 1.

From above all, the HSRC algorithmic procedures can be summarized as Algorithm 1:

4 Experiments and Discussions

In this section, we perform experiments on face databases to demonstrate the efficiency of HSRC and compare the recognition rate with a 95 percent confidence interval. To evaluate more comprehensively the performance of HSRC, in Sect. 4.1 we first test FR without occlusion, and then in Sect. 4.2 we demonstrate the robustness and high efficiency of the proposed

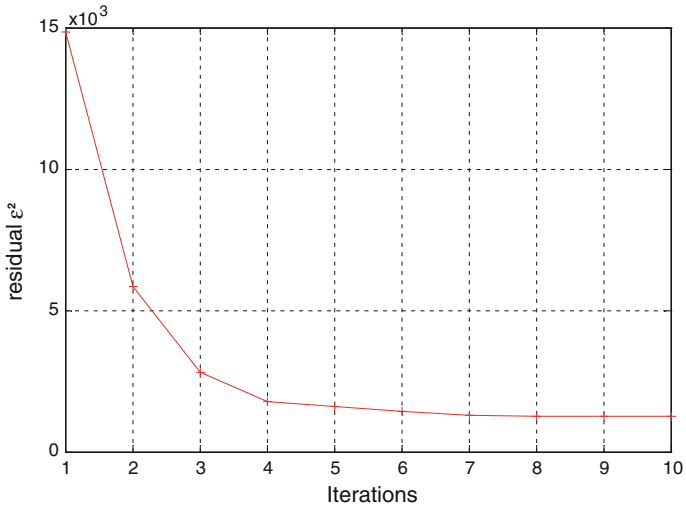

Fig. 1 Illustration of the convergence of HSRC

Algorithm 1 Heteroscedastic sparse representation based classification

Step 1: Input m training samples $A \in \Re^{n \times m}$ partitioned into k classes, and a test sample y

Step 2: Compute initial coefficient $x = (A^T A)^{-1} A^T y$ by ordinary least squares estimation

Step 3: Compute the residual $\varepsilon_{=} y - Ax$

Step 4: Compute the heteroscedastic weights

First heteroscedastic term σ_i can be achieved by Eq. 13, then heteroscedastic weight W can be obtained by:

$W = diag(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_n^{-1})$

Step 5: Compute the coefficient of the model.

$\hat{x} = (A^T W A + \lambda)^{-1} A^T W y$ where λ is positive scalar.

Step 6: Go back to step 3 until the maximum number of iterations is reached or the condition of convergence is met.

Step 7: Compute the residuals

$r_j(y) = \left\| y - A\delta_j(\hat{x}) \right\|_2$, for $j = 1,2,\ldots$ k.

where $\delta_j(x)$ is the characteristic function which selects the coefficients associated with the j-th class.

Step 8: Output that identity(y) = argmin $r_j(y)$.

method to random block occlusion. Finally in Sect. 4.3 we test FR against real face disguise occlusion. To evaluate the proposed HSRC method, we systematically compare it with the SVM (support vector machine), SRC, GSRC (Gabor-SRC) method in real-work database: Extended Yale B [8], AR [20]. In the experiments, we adopted cross-validation strategy for recognition. The data set was splitted into two parts. One part was taken for training, and the other part would be used for testing. The experiments demonstrate good results can be achieved by HSRC when λ is assigned a small positive value from 0.000001 to 0.1. For simplicity, the parameter λ is set as 0.001 by default.

Table 1 Accuracy and running time on extended Yale B database

	SVM	SRC	GSRC	HSRC
Dimensions ($d = 50$)				
Accuracy	93.42%	93.59%	94.78%	**96.22%**
Training time	0.0072	0.0039	0.0983	0.0048
Testing time	0.0002	2.1046	2.7623	0.5233
Total time	0.0074	2.1085	2.8606	0.5281
Dimensions ($d = 100$)				
Accuracy	94.32%	95.07%	96.87%	**97.31%**
Training time	0.0074	0.0040	0.0994	0.0151
Testing time	0.0004	3.1842	3.8668	0.5823
Total time	0.0078	3.1882	3.9962	0.5974
Dimensions ($d = 200$)				
Accuracy	96.64%	97.04%	98.23%	**98.89%**
Training time	0.0076	0.0041	0.1001	0.0359
Testing time	0.0009	8.0362	10.230	0.6771
Total time	0.0085	8.0403	10.3301	0.7130

The bold value indicates the best accuracy rate

4.1 Face Recognition Without Occlusion

(1) The Extended Yale B dataset consists of 2,414 frontal face images of 38 subjects. They are captured under various lighting conditions and cropped and normalized to 192 × 168 pixels. The face images were captured under various illumination conditions. We randomly split the database into two halves. One half (about 32 images per person) was used for training, and the other half for testing. The images are reduced to 50, 100, 200 dimensions, respectively. Here, for computational convenience, the size of image is cropped to 32 × 32. Here we set β= 0.8. Table 1 illustrates the face recognition rates and running times under different methods. We can see that the recognition rates and running times increase with the larger dimensions. Our HSRC method achieves a recognition rate between 96.22% and 98.89%, much better than the other methods. More importantly both training time and testing time are greatly reduced compared with SRC and GSRC.

(2) The AR dataset consists of over 3,000 frontal images of 126 individuals. There are 26 images of each individual, taken at two different occasions. The faces in AR contain variations such as illumination change, expressions and facial disguises. We selected 100 subjects (50 male and 50 female) for out experiments. For each subject, we randomly take the first half for training and the rest for testing. Some AR images are shown as Figure 2. Here, for computational convenience, the size of image is cropped to 33 ×

Fig. 2 All the 14 images of the first person in the subset of AR. The first seven images are from the first session, and the last seven images are from the second session

Table 2 Accuracy and running time on AR database

	SVM	SRC	GSRC	HSRC
Dimensions ($d = 50$)				
Accuracy	68.57%	82%	83.14%	**86.05%**
Training time	0.0071	0.0083	0.2492	0.0098
Testing time	0.0002	0.9380	1.8065	0.5710
Total time	0.0073	0.9463	2.0557	0.5808
Dimensions ($d = 100$)				
Accuracy	72.57%	89%	91.29%	**93.43%**
Training time	0.0073	0.0086	0.2469	0.0301
Testing time	0.0002	0.9720	2.4001	0.6388
Total time	0.0075	0.9806	2.6470	0.6689
Dimensions ($d = 200$)				
Accuracy	74.71%	92.29%	93.43%	**95.81%**
Training time	0.0079	0.0094	0.2427	0.0723
Testing time	0.0004	1.2989	3.1371	0.7081
Total time	0.0083	1.3083	3.3798	0.7804

The bold value indicates the best accuracy rate

24. Here we setβ= 0.9. Table 2 shows the recognition rates for this experiment. HSRC achieves accuracy rate of 95.81% with 200 dimensional feathers, higher than the other methods, e.g., 74.71% for SVM, 92.29% for SRC, 93.43% for SRC, while the running times of HSRC have great advantage over the other sparse representation method (Table 3).

4.2 Face Recognition with Block Occlusion

The next one is more challenging, we test the efficiency of HSRC to the block occlusion using the Extended Yale B face dataset. We randomly take the first half for training and the rest for testing. We simulate various levels of contiguous occlusion, from 30% to 50%, by replacing a randomly located square block of each test image with an unrelated image, Again, the location of occlusion is randomly chosen for each image and is unknown to the computer. A test example of Extended Yale B with 30% occluded block is shown as Figure 3. Here, for computational convenience, the size of image is cropped to 32×32. Here we set β= 0.8. The results of the experiments are more exciting, and listed in Table 3. The accuracy rate of all the methods decline with the occlusion levels increasing, which indicates that loss of feature affects the face recognition performance. But HSRC preserves good performance of 63.92% when the occlusion percentage is 50%. Moreover HSRC is significantly about 9 times faster than SRC, while more than about 33% improvement in recognition rate. The results lie in the facts that SRC need a large occlude dictionary, while HSRC does not need one.

4.3 Face Recognition with Disguise

A subset from the AR database consists of 1399 images from 100 subjects, 50 male and 50 female. For training, we use 799 images (about 8 per subject) of unoccluded frontal views with varying facial expression, and we consider two separate test sets of 200 images. For

Table 3 Accuracy and running time under different levels of block occlusion

	SRC	GSRC	HSRC
Occlusion (30%)			
Accuracy	81.5%	82.7%	**83.34%**
Training time	0.0269	0.1002	0.0131
Testing time	63.5033	11.8324	0.7201
Total time	63.5302	11.9326	0.7332
Occlusion (40%)			
Accuracy	60.61%	65.41%	**66.72%**
Training time	0.0279	0.1011	0.0129
Testing time	63.3135	10.9461	0.7122
Total time	63.3414	11.0472	0.7252
Occlusion (50%)			
Accuracy	47.86%	59.81%	**63.92%**
Training time	0.0283	0.1015	0.0127
Testing time	63.1444	10.9355	0.6901
Total time	63.1727	11.0370	0.7028

The bold value indicates the best accuracy rate

Fig. 3 An test example of Extended Yale B with 30% occluded block

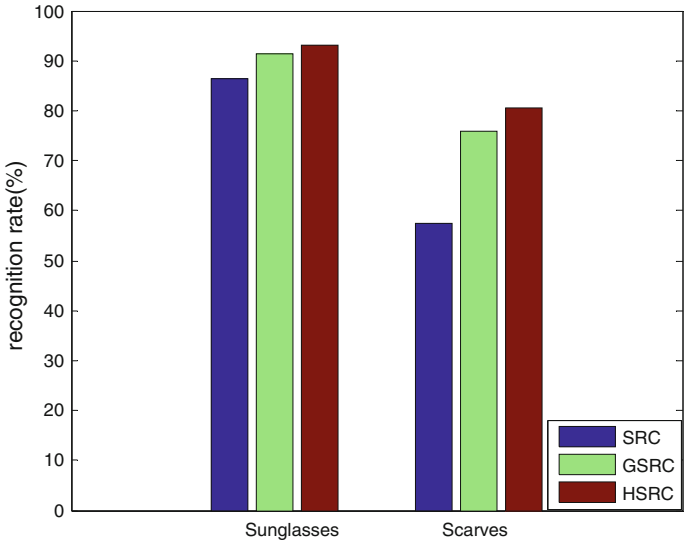

Fig. 4 Accuracy on disguise AR database

testing, we consider two separate test sets of 200 images. The first test set contains images of the subjects wearing sunglasses, which occlude roughly 20% of the image. The second test set contains images of the subjects wearing a scarf, which occludes roughly 40% of the image. For computational convenience, all the images were resized to 33×24. Here we set $\beta = 0.9$.

Figure 4 shows the face recognition results of three different methods under sunglasses and scarves. The test set with sunglasses only occlude roughly 20% of the image, so all the methods achieve good performance. Especially about 8% improvement is obtained by HSRC. In the test set with scarves, all the method decline greatly, but HSRC still preserve 80.5% recognition rate, which demonstrate our method is more robust than other methods.

5 Conclusion

This paper proposed a novel method, called Heteroscedastic sparse representation based classification (HSRC). The HSRC method is attractive due to its robustness and its high efficiency. On the one hand, HSRC is robust to various types outliers (i.e., occlusion, corruption, expression, etc.) because heteroscedastic correction have been adopted for homoscedasticity by weighting various residual with heteroscedastic estimation. On the other hand, HSRC reduces greatly running times owing to avoiding a large occlude dictionary. The experimental results on Extended Yale B and AR database show that HSRC consistently outperforms the classical methods.

Acknowledgements This work is partially supported by the National Science Foundation of China under grant No. 60973098

References

1. Aiazzi B, Alparone L, Baronti S (1999) Estimation based on entropy matching for generalized Gaussian pdf modeling. IEEE Signal Process Lett 6(6):138–140
2. Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. SIAM J Sci Comput 20(1):33–61
3. Chen SS, Donoho DL, Saunders MA (1999) Atomic decomposition by basis pursuit. SIAM J Sci Comput 20(1):33–61
4. Chung F (1997) Spectral graph theory. Regional conference series in mathematics, no. 92
5. Donoho D (2006) Compressed sensing. IEEE Trans Inform Theory 52(4):1289–1306
6. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15(12):3736–3745
7. Francois M, Pierre V, Jean-Philippe T (2006) Matching pursuit-based shape representation and recognition using scale-space. Int J Imaging Syst Technol 6(5):162–180
8. Georghiades A, Belhumeur P, Kriegman D (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans Pattern Anal Mach Intell 23(6):643–660
9. Gragg JG (1983) More efficient estimation in the presence of heteroscedasticity of unknown form. Econometrica 51(3):751–764
10. Gribonval R, Vandergheynst P (2006) On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. IEEE Trans Inform Theory 52(1):255–261
11. He X, Niyogi P (2003) Locality preserving projections. In: Proceedings of 16th conference neural information processing systems
12. Huang J, Yuen PC, Chen W (2007) Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations. IEEE Trans Syst Man Cybern B Cybern 37:847–862
13. Joliffe I (1986) Principal component analysis. Springer, New York
14. Jonathon Phillips P (1998) Matching pursuit filters applied to face identification. IEEE Trans Image Process 7(8):1150–1164

15. Kong H, Wang L, Teoh EK (2005) A framework of 2D fisher discriminant analysis: application to face recognition with small number of training samples. Proc IEEE Int Conf Comput Vis Pattern Recognit 2:1083–1088
16. Li H, Jiang T, Zhang K (2004) Efficient and robust feature extraction by maximum margin criterion. In: Proceedings of the advances in neural information processing systems
17. Long JS, Ervin LH (2000) Using heteroscedasity consistent stand errors in the linear regression model. Am Stat 54:217–224
18. Mairal J, Elad M, Sapiro G (2008) Sparse representation for color image restoration. IEEE Trans Image Process 17(1):53–69
19. Mallat S, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 41(12):3397–3415
20. Martinez A, Benavente R (1998) The AR face database. CVC Tech. Report No. 24
21. Meignen S, Meignen H (2006) On the modeling of small sample distributions with generalized Gaussian density in a maximum likelihood framework. IEEE Trans Image Process 15(6):1647–1652
22. Muller HG, Stadtmuller U (1987) Estimation of heteroscedasticity in regression analysis. Ann Stat 15(2):610–625
23. Murray J, Kreutz-Delgado K (2007) Visual recognition and inference using dynamic overcomplete sparse learning. Neural Comput 19:2301–2352
24. Patel VM, Nasrabadi NM, Chellappa R (2010) Automatic target recognition based on simultaneous sparse representation. In: International Conference on Image Processing, pp 1377–1380
25. Rigamonti R, Brown M, Lepetit V (2011) Are sparse representations really relevant for image classification? In: Computer vision and pattern recognition, pp 1545–1552
26. Sharifi K (1995) Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. IEEE Trans Circuits Syst Video Technol 5:52–56
27. Shi Q, Eriksson A, Hengel A (2011) Is face recognition really a compressive sensing problem? In: Computer vision and pattern recognition, pp 553–560
28. Tropp JA (2004) Greed is good: algorithmic results for sparse approximation. IEEE Trans Inform Theory 50(10):2231–2242
29. Varanasi MK, Aazhang B (1989) Parametric generalized Gaussian density estimation. J Acoust Soc Am 86:1404–1415
30. Wright J, Yang A, Ma Y (2009) Robust face recognition via sparse representation. IEEE T-PAMI 31(2):210–227
31. Yang M, Zhang L (2010) Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In: Europeon conference on computer vision
32. Yang A, Wright J, Ma Y (2007) Feature selection in face recognition: a sparse representation perspective. UC Berkeley Technical Report UCB/EECS-2007-99
33. Zhang L (2011) Sparse representation or collaborative representation which helps face. In: International conference on computer vision