Cost-Sensitive Sparsity Preserving Projections for Face Recognition

Xiaoyuan Jing[1,2] , Wenqian Li[1]*, Hao Gao[1], Yongfang Yao[1], Jiangyue Man[1]
[1]College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China
[2]State Key Laboratory of Software Engineering, Wuhan University, Wuhan, China
*Email: joycelee19871228@hotmail.com

***Abstract*--As one of the most popular research topics, sparse representation (SR) technique has been successfully employed to solve face recognition task. Though current SR based methods prove to achieve high classification accuracy, they implicitly assume that the losses of all misclassifications are the same. However, in many real-world face recognition applications, this assumption may not hold as different misclassifications could lead to different losses. Driven by this concern, in this paper, we propose a cost-sensitive sparsity preserving projections (CSSPP) for face recognition. CSSPP considers the cost information of sparse representation while calculating the sparse structure of the training set. Then, CSSPP employs the sparsity preserving projection method to achieve the projection transform and keeps the sparse structure in the low-dimensional space. Experimental results on the public AR and FRGC face databases are presented to demonstrate that both of the proposed approaches can achieve high recognition rate and low misclassification loss, which validate the efficacy of the proposed approach.**

Keywords-cost-sensitive learning; sparse representation; cost-sensitive classifier; feature extraction; face recognition

I. INTRODUCTION

Face recognition is one of the most popular and challenging research topics in image processing and pattern recognition [1]. Since face images have a high dimensionality, feature extraction methods play a very important role in reducing the dimension before classification. Some typical feature extraction methods include PCA [2] and LDA [3]. PCA performs feature extraction by projecting the original high-dimensional face images to a low-dimensional linear subspace spanned by the leading eigenvectors of a covariance matrix. LDA seeks a linear transformation that maximizes the between class scatter and simultaneously minimizes the within class scatter.

In recent years, while data variety is important to be concerned, many other feature extraction methods study manifold structure [4], for example, locality preserving projections (LPP) [5] and Sparsity Preserving Projections(SPP) [6]. LPP finds an embedding that preserves information, and obtains a face subspace that best detects the essential face manifold structure. Unlike LPP, where local neighborhood information is preserved during dimensionality reduction procedure, SPP aims to preserve the sparse reconstructive relationship of the data, which is achieved by minimizing a l^1 regularization-related objective function [7].

However, all the feature extraction methods we have mentioned above aim to attain low recognition errors and assume same loss from all misclassifications. In the real world face recognition applications, this assumption may not hold since different misclassifications could lead to different losses. Zhang et al. [8] presented a multiclass cost-sensitive learning framework for face recognition which aims at minimizing the total loss of misclassifications instead of classification errors. Further, he proposed two cost-sensitive classifier based methods, namely multiclass cost-sensitive KLR (mcKLR) and multiclass cost-sensitive *k*-nearest neighbor (mckNN). Lu et al.[9] incorporates a cost matrix, which specifies the different costs associated with misclassifications of subjects, into three popular subspace learning algorithms and the corresponding cost-sensitive methods, namely CSPCA, CSLDA and CSLPP.

Inspired by Lu's work, we propose a cost-sensitive sparsity preserving projections (CSSPP) for face recognition. CSSPP considers the cost information of sparse representation while calculating the sparse structure, which is called the l^1 graph, of the training set. Then, CSSPP employs the sparsity preserving projections method to achieve the projection transform and keep the sparse structure in the low-dimensional space. In addition, we further analyze the relationship between costs and the sparse coefficients. Experiments are conducted on AR and FRGC face databases, and results prove that CSSPP achieves a minimum overall recognition loss by performing recognition in the low-dimensional subspaces derived.

The rest of the paper is organized as follows: In section 2, we analyze the relationship between costs and the sparse coefficients. In section 3, our proposed approach, CSSPP, is presented. Section 4 presents the algorithm procedures of CSSPP. Experiments are described in section 5 and section 6 concludes the whole paper.

II. RELATIONSHIP BETWEEN COSTS AND SPARSE COEFFICIENTS

Let $A=[x_1, x_2, \cdots, x_N] \in R^{n\times N}$ be the set of training samples. *A* contains the elements of an overcomplete dictionary in its columns, that is, a sample $y \in R^n$ can be represented using a few entries of *A*. The sparse coefficients $\tilde{X}$ can be obtained by solving the following minimization problem:

$$\tilde{X} = \arg\min_X \|y - AX\|_2 + \lambda \|X\|_1 . \qquad (1)$$

Obviously, $\tilde{X}$ is determined by the sample set *A* and the test sample *y*. However, in some cases, $\tilde{X}$ seems to be idealization that we assume same loss from all misclassifications. This assumption does not hold in real world. Driven by this concern, it is considered to obtain a group of cost-sensitive coefficients. Some of the coefficients which correspond to some specific classes should be small.

Firstly, we newly design the costs in sparse representation. Denote a face image as *y*, and its class label as $\phi^*(y)$, respectively. Given *M* gallery subjects with their class labels $\{G_i\}_{i=1,\ldots,M}$ and many impostors, which are regarded as of a metaclass with label *I*. Three types of costs in sparse representation are newly designed as follows:

1) False acceptance representation C_{IG} : Via linear combination, an imposter sample is represented by some gallery samples of the overcomplete sample set.

2) False rejection representation C_{GI} : Via linear combination, a gallery sample is represented by some imposter samples of the overcomplete sample set.

3) Representation by using the same type of samples C_{GG} or C_{II} : Via linear combination, a gallery sample is represented by some other gallery samples of the overcomplete sample set, or an imposter sample is represented by some other imposter samples of the overcomplete sample set.

Driven by the analysis, sparse representation can be improved as follows:

1) Representation of a gallery sample y_G : y_G is considered to be represented by as many gallery samples as possible, while impostor samples of the overcomplete sample set are deemphasized. Thus, the sparse coefficients will not only well reflect the category information of y_G, but also take the costs information of false rejection into consideration. Under this circumstance, two types of costs, C_{GI} and C_{GG}, are used.

2) Representation of an impostor sample y_I : Contrary to the first representation, in hope of applying as many impostor as possible to represent y_I, while ignoring the gallery samples, C_{IG} and C_{II} are the two types of costs we mainly consider. y_I represented by gallery samples may cause false acceptance in classification phase. False acceptance could result in much more serious loss than false rejection. From the view of quantitative, the value of C_{IG} is higher than C_{GI}.

Based on the theory above, we improve (1) by applying a newly defined sample set $\hat{A}$, which contains costs information. The modified formula is as follows:

$$\tilde{X} = \arg\min_X \|y - \hat{A}X\|_2 + \lambda \|X\|_1 . \qquad (2)$$

The element of $\hat{A}$ is calculated as:

$$\hat{x}_i = cost(x_i) \cdot x_i , \qquad (3)$$

where $cost(x_i)$ is the cost function of x_i in sparse representation. According to the two situations of the sparse presentation we have discussed above, $cost(x_i)$ is conducted precisely as follows:

1） Representation of a gallery sample y_G **:**

$$cost(x_i) = \begin{cases} C_{GG}, & \text{if } x_i \text{ is a gallery sample,} \\ C_{GI}, & \text{if } x_i \text{ is an impostor sample.} \end{cases} , \qquad (4)$$

2) Representation of an impostor sample y_I :

$$cost(x_i) = \begin{cases} C_{IG}, & \text{if } x_i \text{ is a gallery sample,} \\ C_{II}, & \text{if } x_i \text{ is an impostor sample.} \end{cases} , \qquad (5)$$

where C_{GI} , C_{IG} , C_{GG} and C_{II} usually satisfy the inequality $C_{IG} > C_{GI} > C_{GG} = C_{II}$. Here, we set $C_{IG}=20$, $C_{GI}=2$, $C_{GG} = C_{II} = 1$.

In (2), the overcomplete sample set *A* is weighted by a cost-sensitive function, which is the common way in improving the cost-sensitive learning methods. On the premise that $\|y - \hat{A}X\|_2$ in (2) is fixed, if the value of $\hat{x}_i$ increases by applying the weighted cost-sensitive function, the corresponding sparse coefficients would be decrease obviously. In other words, if the cost value, which is the weight, in sparse representation is large, the sparse coefficients will become small.

III. COST-SENSITIVE SPARSITY PRESERVING PROJECTIONS

Given a test sample in a real application, since the category information is uncertain, the type of the cost-sensitive sparse representation we have analyzed in Section Ⅱ can not be figured out. In training phase of Sparsity Preserving Projections (SPP), categories of the training samples are open to us; hence the sparse coefficients of each training sample are easily obtained apparently. Inspired by SPP, we proposed a cost-sensitive learning based approach namely cost-sensitive sparsity preserving projections (CSSPP).

Let $A = [x_1, x_2, \cdots, x_N] \in R^{n \times N}$ be the set of training samples, where x_i is a *n*-dimensional vector. $\hat{A}$ is the weighted training set which can be calculated by (3), (4) and (5). Similar to (2), the cost-sensitive sparse coefficients $s_i \in R^N$ can be obtained by solving the following l^1 minimization problem:

$$s_i = \arg\min_{s_i} \left\| x_i - \hat{A}s_i \right\|_2 + \lambda \left\| s_i \right\|_1 . \quad (6)$$

$s_i = [\alpha_{i,1}, \cdots, \alpha_{i,i-1}, 0, \alpha_{i,i+1}, \cdots, \alpha_{i,N}]^T$, the i^{th} element $\alpha_{i,i}$ of s_i is set as zero, which means the given sample x_i can be reconstructed by some other training samples except by using itself. All the coefficients are affected by the cost weight. (6) is a l^1 minimization problem, which can be solved by standard linear programming[10].

For CSSPP, we construct the following objective function to seek the projections which best preserve the optimal cost-sensitive coefficient matrix:

$$\min_W \sum_{i=1}^{N} \left\| W^T x_i - \sum_{j=1}^{N} \alpha_{i,j} W^T x_j \right\|^2 . \quad (7)$$
$$s.t.\ W^T AA^T W = I$$

In particular, here we use the original training sample set A instead of $\hat{A}$. The weighted sample set $\hat{A}$ is already applied in calculating the cost-sensitive sparse coefficients. The matrix of the coefficients we have obtained is also called a cost-sensitive l^1 graph. In l^1 graph, structure of the original samples are actually depends on the sparse coefficients, and then the low-dimensional embedding of data is evaluated to best preserve such cost-sensitive structure in the original high-dimensional data space. Therefore, as the original set, A is adopted here. The optimization problem can be reformulated into an eigen-problem:

$$AMA^T W = \lambda AA^T W , \quad (8)$$

where $M = (I-S)^T (I-S)$, $S = [s_1, s_2, \cdots, s_N]$. The optimal solution W is composed by the eigenvectors of matrix $(AA^T)^{-1} AMA^T$ associated with the *d* largest eigenvalues.

The projected *d*-dimensional data set Z is obtained by $z = W^T A$.

IV. CSSPP ALGORITHM

Based on the above discussion, we summarize the proposed algorithm as follows:

Algorithm: cost-sensitive sparsity preserving projections

Step 1: Construct the cost-sensitive coefficient matrix S by solving the l^1 minimization problem of (6).

Step 2: Calculate the projection vectors using (8), and the eigenvectors of $(AA^T)^{-1} AMA^T$ corresponding to the largest *d* eigenvalues span the optimal subspace.

Step 3: The projected *d*-dimensional data set Z is obtained by $z = W^T A$.

V. EXPERIMENTS

A. Database Introduction

In our experiments, we use the public AR [11] and FRGC [12] databases to validate the effectiveness of the proposed approach.

AR database contains more than 4000 color pictures of 126 people, including frontal views of faces with different facial expressions, various occlusions and under different lighting conditions. Our experiments will utilize the 3094 (=119*26) samples of the 119 people. Each image contains 256 grey levels. We will minimize the image into 60*60. All the cropped images of one subject are shown in Fig. 1.

The face recognition grand challenge version 2 experiment 4 (shorten to FRGC-v2-e4) contains 12776 training images, 16028 controlled target images and 8014 uncontrolled query images. The training image set contains 222 individuals, each 36-64 images. We choose 36 images of each individual and then crop every image to the size of 60*60. All images of one subject are shown in Fig. 2.

For each database, the train sample set in our experiments includes 7 images with 20 gallery subjects and 7 images with 10 imposter subjects randomly selected from the entire database. All the impostor subjects are considered as a metaclass. The rest samples corresponding to each chosen subject are used as the testing sample set. We repeat this selection 10 times for each database and calculate the average recognition results.

We compare our approach with PCA [2], CSPCA [9], LPP [5], CSLPP [9] and SPP [13]. The parameters C_{GI}, C_{IG} and C_{GG} are specified as 20:2:1.

B. Experimental results

We compare the total cost, total error rate (err), error rate of false acceptance (err_{IG}), and error rate of false rejection (err_{GI}) of all the compared methods. The results are shown in Table 1 and Table 2. Equations of total cost, err, err_{IG} and err_{GI} are listed as follows:

$$\begin{aligned}\text{Total cost} = C_{IG} \bullet \text{Number of false acceptance} + \\ C_{GI} \bullet \text{Number of false rejection} + \\ C_{GG} \bullet \text{Number of false identification}\end{aligned} , \quad (11)$$

$$\text{err} = \frac{\text{Error number}}{\text{Total testing number}} \times 100\% , \quad (12)$$

$$\text{err}_{IG} = \frac{\text{Number of false acceptance}}{\text{Total error number}} \times 100\% , \quad (13)$$

$$\text{err}_{GI} = \frac{\text{Number of false rejection}}{\text{Total error number}} \times 100\% . \quad (14)$$

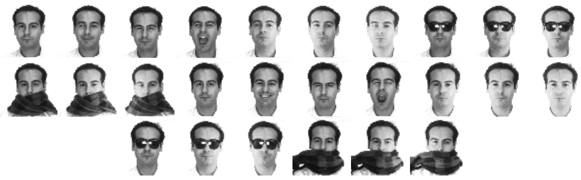

Figure 1. Demo images of one subject from the AR face database

Figure 2. Demo images of one subject from the FRGC face database

TABLE I. TOTAL COST, err , err_{IG} AND err_{GI} ON AR DATABASE

Methods	**Cost (100%)**	err_{IG} **(100%)**	err_{GI} **(100%)**	err **(100%)**
PCA	444.8	26.34	22.85	12.49
CSPCA	400.8	21.21	27.34	13.33
LPP	393	24.62	28.68	11.61
CSLPP	310.4	7.52	65.71	16.95
SPP	392.8	29.17	26.84	10.18
CSSPP	**297.4**	20.91	40.40	**9.61**

TABLE II. TOTAL COST, err , err_{IG} AND err_{GI} ON FRGC DATABASE

Methods	**Cost (100%)**	err_{IG} **(100%)**	err_{GI} **(100%)**	err **(100%)**
PCA	835.6	27.27	23.81	23.05
CSPCA	754.4	22.90	27.04	23.82
LPP	778.2	25.96	29.94	21.89
CSLPP	692.2	15.33	56.73	27.26
SPP	760.8	26.53	28.00	20.95
CSSPP	**649.6**	23.17	30.15	**19.96**

From Table 1 and Table 2, we can find that the cost-sensitive subspace learning methods (CSPCA, CSLPP, CSSPP) have a smaller total cost than the existing, uniform-cost subspace learning methods (PCA, LPP, SPP). It's notable that compared with the normal subspace learning methods, the false acceptance rates drop while the false rejection rates rise in cost-sensitive methods, since false acceptance may cause much more cost than false rejection, the total costs drop ultimately. In addition, the total error rates of the proposed methods CSSPP on two databases are lower than that of SPP, while the phenomenon comes up in an opposite side in CSPCA and CSLPP. Because CSSPP is modified as a supervised approach while SPP is unsupervised. On the other hand, cost-sensitive coefficients are superior to the normal coefficients in avoiding false representation. Hence the proposed CSSPP achieves the best performance among all the compared methods.

VI. COPYRIGHT FORMS AND REPRINT ORDERS

In this paper, we improve traditional sparsity preserving projections (SPP) by incorporates cost information into sparse representation. We name the proposed method cost-sensitive sparsity preserving projections (CSSPP). CSSPP considers the cost information of sparse representation while calculating the sparse structure of the training set. Then, CSSPP employs the sparsity preserving projections method to achieve the projection transform and keeps the sparse structure in the low-dimensional space. Experimental results on two face databases show the effectiveness of out proposed approach.

ACKNOWLEDGMENT

The work described in this paper was fully supported by the NSFC under Project No. 61073113 and Project No. 61272273, the New Century Excellent Talents of Education Ministry under Project No. NCET-09-0162, the Doctoral Foundation of Education Ministry under Project No. 20093223110001.

REFERENCES

[1] D. Marr, "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information," W.H. Freeman, San Francisco, 1982.

[2] M. Turk and A. Pentland, "Eigenfaces for Recognition," J. Cognitive Neuroscience, vol. 3, no. 1, 1991.

[3] P. N. Belhumeur , J. P. Hespanha and D. J. Kriengman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp.711 -720, 1997.

[4] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, pp.2323 -2326, 2000.

[5] X. He , S. Yan , Y. Hu , P. Niyogi and H. Zhang, "Face Recognition Using Laplacianfaces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, pp.328 -340, 2005.

[6] L. Qiao , S. Chen and X. Tan, "Sparsity preserving projections with applications to face recognition," Pattern Recognition, vol. 43, pp.331 -341, 2010.

[7] B. Cheng, J. Yang, S. Yan, Y. Fu and T. Huang, "Learning With L1-Graph for Image Analysis," IEEE Trans. Image Processing, vol. 19, pp. 858-866, 2010.

[8] Y. Zhang and Z. H. Zhou, "Cost-sensitive face recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, 2010.

[9] J. Lu and Y. Tan, "Cost-sensitive subspace learning for face recognition," IEEE Conf. Computer Vision and Pattern Recognition, pp. 2661-2666, 2010.

[10] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM J. Sci Comp., vol. 20, pp.33 -61, 1999.

[11] A. M. Martinez and R. Benavente, "The AR Face Database," 1998.

[12] P. J. Phillips , P. J. Flynn , T. Scruggs , K. W. Bowyer , J. Chang , K. Hoffman , J. Marques , J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, pp.947 -954, 2005.

[13] J. Wright , A. Yang , A. Ganesh , S. Sastry and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, pp.210 -227, 2009.