Evaluation of Distance Measures For NMF-Based Face Image Applications

Yun Xue[a], Chong Sze Tong[b], Tiechen Li[a]
[a] School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou Guangdong 510631, China
Email: xueyun@scnu.edu.cn, ltch2013@gmail.com
[b] Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
Email: cstong@math.hkbu.edu.hk

***Abstract*— Non-negative matrix factorization (NMF) is an increasingly popular feature extraction method. Since it is designed to fit training samples using linear combination of non-negative basis vectors, it is particular suitable for image applications as it affords intuitive localized interpretations. However, in this space defined by the NMF basis images, there has not been any systematic research to identify suitable distance measure for NMF-based data classification.**

In this paper, the performance of 19 distance measures between feature vectors is evaluated based on the result of the NMF algorithm for face recognition, which include most of the standard distance measures used in face recognition, as well as two new non-negative vector similarity coefficient-based (NVSC) distances that we recommend for use in NMF-based pattern recognition. Recognition experiments are performed using the CMU AMP Face Expression database, CBCL2 database, MIT-CBCL database, YaleB database, and FERET database.

We also compared the performance of NMF with Eigenface method and showed that the NMF algorithm using the NVSC distance yielded the best recognition results.

***Index Terms*— face recognition, non-negative matrix factorization, distance measures**

I. Introduction

During the past decade, face recognition has attracted significant attention for its wide range of applications. Principal Component Analysis (PCA) (i.e. Eigenface) has been proven to be a successful face-based method to this problem [1].

Whereas, it has some limitations. Firstly, although PCA gives a very good representation of the images, it has a poor discriminatory ability. Secondly, PCA basis images lack intuitive visual meaning. Finally, the case with occlusions is difficult to deal with since it is based on extracting global face features.

Recently, a new method has been proposed for obtaining a linear representation of data [2], which is called non-negative matrix factorization (NMF), differing from other methods by the usage of non-negativity constraints. The initial data matrix representing the whole database is approximately factorized into two nonnegative matrix factors and consequently a part based representation of images are produced because it allows only additive, not subtractive, combinations of basis images.

In the paper, all the faces are projected into this NMF space to obtain their corresponding feature vectors, and then the distance between these vectors is calculated. Although there exist many distance, such as the Euclidean distance, the L1 and Mahalanobis distance, we are able to find only few attempts to propose, compare and use other distance measures [3]–[6] for NMF-based face recognition to achieve better recognition results. In [3], [4], a few distance measures are discussed: the L1, L2, Cosine and Earth Movers Distance (EMD), Cosine and Earth Movers Distance (EMD), and a handwritten digit database (MNIST) is used to obtain the performance evaluation of different distance measures. There was no relevant result corresponding to face image databases.

In this paper, the recognition performance of 19 distance measures are compared, including two new non-negative vector similarity coefficient-based (NVSC) distance measures that we are advocating for use in NMF-based face recognition. The experiments show that these new distance measures are always among the best distance measures with respect to different image databases and at different settings.

We have used the Principal Component Analysis (i.e. Eigenface) combined with its some distance measures in common use for a direct comparison, and the experimental result also supports the conclusion that our new distance measure combined with NMF can achieve a better performance when identifying the probe images in database.

This paper is organized as follows. The background theory of PCA and NMF are introduced in Section 2. Sect.3 introduced the detailed definition of distance measures used in this paper. In Sect.4, some description of the image databases used in the paper are given. Sect.5 discusses the experimental results of our face recognition system based on the NMF algorithm. Finally, the conclusions and the future work are presented in Sect.6.

II. Review of PCA and NMF

This section provides the background theory of PCA and NMF for face recognition, which are both unsupervised learning methods.

doi:10.4304/jcp.9.7.1704-1711

A. Principal Component Analysis

Let $X = \{X_n \in R^d | \ n = 1, \dots, N\}$ be an set of vectors. First, the mean vector EX and covariance matrix M are computed for the full data set.

$$EX = \frac{1}{N}\sum_{n=1}^{N} X_n,$$

$$\hat{X} = \{\hat{X}_n, n = 1, \dots, N\} \quad \text{with} \quad \hat{X}_n = X_n - EX,$$

$$M = cov(\hat{X}),$$

with

$$M(i,j) = \frac{1}{N-1}\sum_{n=1}^{N}(\hat{X}_n(i)\hat{X}_n(j)), 1 \le i, j \le d.$$

Next, the eigenvectors and eigenvalues of the matrix M are computed, and sorted according to decreasing eigenvalue. From matrix theory, it's well known the eigenvectors of the matrix M form an orthonormal basis.

Finally, the largest k such eigenvectors are chosen, then the PCA of a vector y can be calculated by projecting it onto the subspace which is spanned by these k eigenvectors. It can be shown that this representation minimizes a squared error criterion [7].

B. Eigenfaces

Essentially, Eigenface is the eigenvector associated with large eigenvalues from the PCA method. After representing a face image using a weighted sum of eigenfaces, face recognition is performed by comparing the corresponding weight vectors between probe and reference faces.

C. NMF method

The algorithm is to acquire a linear representation of data under non-negativity constraints. The following is the basic idea.

First, an image database is represented as a $n \times m$ matrix V, with each column corresponding to a initial face image, including n non-negative elements characterizing the pixel value and m is the number of images.

Then two new non-negative matrices (W and H) are found to approximate the original matrix [2].

$$V_{ij} \simeq (WH)_{ij} = \sum_{a=1}^{r} W_{ia}H_{aj}, W \in R^{n\times r}, H \in R^{r\times m}. \tag{1}$$

where matrix W consists of r non-negative basis vectors and column vectors of H mean the weights when approximating the corresponding column in V using the bases from W.

No subtractions occur in the above NMF procedure compared with the PCA approach, thus the non-negativity constraints are compatible with the intuitive idea of combining parts to form a whole face.

The update rule for NMF is shown below:

First an objective function to characterize the similarity between V and WH is constructed:

$$F = \sum_{i=1}^{n}\sum_{j=1}^{m}\left[V_{ij}\log\frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}\right]. \tag{2}$$

Then an iterative algorithm converging to a local maximum of this objective function is derived [2]:

$$W_{ia} \leftarrow W_{ia}\sum_{j}\frac{V_{ij}}{(WH)_{ij}}H_{aj}, \tag{3}$$

$$W_{ia} \leftarrow \frac{W_{ia}}{\sum\limits_{j} W_{ja}}, \tag{4}$$

$$H_{aj} \leftarrow H_{aj}\sum_{i} W_{ia}\frac{V_{ij}}{(WH)_{ij}}. \tag{5}$$

For the application of face recognition, the NMF algorithm includes training and recognition stages, which are detailed as below.

D. NMF-based training stage

This process contains 3 major steps. In the first step, a $n \times m$ matrix V_1 is used to represent all the training images in one database.

Secondly, the NMF algorithm is applied to V_1 and two new matrices (W_1 and H_1) are obtained as in Sect.II-C, s.t.

$$(V_1)_{ij} \simeq (W_1H_1)_{ij} = \sum_{a=1}^{r}(W_1)_{ia}(H_1)_{aj}$$

where W_1 is the base matrix, and H_1 is the weight matrix.

Finally, different libraries are built to save the training image representations and their corresponding representational bases for all the face databases described in Sect.4.

E. NMF-based recognition stage

Face recognition in the NMF linear subspace is performed as follows.

1) Feature extraction: There are two ways to obtain the feature vectors of training images and test images [8], [9] (in the following text, we will refer to these two approaches as $proj = 1$ or 2).

I. Approach 1 (i.e. $proj = 1$): Let $W^+ = (W_1^T W_1)^{-1} W_1^T$, then we project each training face image V_i into the linear space as a feature vector $H_i' = W^+V_i$ which is then used as a prototype feature point. A test face image V_t to be classified is represented as $H_t' = W^+V_t$.

II. Approach 2 (i.e. $proj = 2$): Using the bases W_1 obtained from the training process, the iterative technique can be used in the original NMF algorithm while keeping W_1 fixed (i.e. do not use the iterative update rules (3) and (4) to update W_1). Then, the weight matrix H_2 could be acquired by using the fixed set of bases (W_1). The matrices H_1 and H_2 are the feature vectors of training images and test images respectively.

2) Classification: In this step, the mean feature vector H_m of each class in the training set is calculated at first; then we calculate all the distance measures (defined in Sect.3) between the feature vector of test image and the mean vector, dist(H_t, H_m); finally, we classify the test image into the class which the closest mean vector belongs to.

III. Distance Measures

Provided that X, Y are feature vectors of length n obtained by NMF method where X is the weight of probe images, and Y is the weight of training images. While σ is the auto-covariance matrix for weight vector of training images, and $\{s_i, i = 1, \cdots, n\}$ represents the square root of diagonal element in σ, i.e. the standard deviation for training images. Then the distances between these feature vectors can be calculated. All the definitions of distance measures used in this paper are described as following [3], [7], [10]–[14].

(1) Manhattan distance (L1 metric, city block distance)

$$d(X,Y) = \sum_{i=1}^{n} |x_i - y_i| \tag{6}$$

(2) Euclidean distance (L2 metric)

$$d(X,Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{7}$$

(3) Chebychev distance (L-∞ norm)

$$d(X,Y) = \max_{1 \le i \le n} |x_i - y_i| \tag{8}$$

(4) Mahalanobis distance

$$d(X,Y) = \sqrt{(X-Y)'\sigma^{-1}(X-Y)} \tag{9}$$

(5) Lance distance

$$d(X,Y) = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|} \tag{10}$$

(6) Statistical distance

$$d(X,Y) = \sum_{i=1}^{n} \left| \frac{x_i - y_i}{s_i} \right| \tag{11}$$

(7) Divergence

$$d(X,Y) = \sum_{i=1}^{n} (x_i \ln \frac{x_i}{y_i} - x_i + y_i) \tag{12}$$

Like the Euclidean distance, it is also bounded below by zero, and vanishes if and only if $X = Y$. But it is not a metric, because it is not symmetric in X and Y, we will refer to it as the divergence of X from Y.

(8) Kullback-Leibler distance (Relative Entropy)

$$d(X,Y) = \sum_{i=1}^{n} x'_i \log_2 \frac{x'_i}{y'_i}, \;\; x'_i = \frac{|x_i|}{\sum\limits_{i=1}^{n} |x_i|}, y'_i = \frac{|y_i|}{\sum\limits_{i=1}^{n} |y_i|} \tag{13}$$

Like divergence, it also is not a metric, because it is not symmetric in X and Y. Symmetrized versions of these two distance measures are given below.

(9) Symmetrized divergence

$$\begin{aligned} d(X,Y) &= \sum_{i=1}^{n} \left(x_i \ln \frac{x_i}{y_i} - x_i + y_i \right) \\ &\quad + \sum_{i=1}^{n} \left(y_i \ln \frac{y_i}{x_i} - y_i + x_i \right) \\ &= \sum_{i=1}^{n} \left(x_i \ln \frac{x_i}{y_i} + y_i \ln \frac{y_i}{x_i} \right) \end{aligned} \tag{14}$$

(10) Symmetrized Kullback-Leibler distance

$$\begin{aligned} d(X,Y) &= \sum_{i=1}^{n} x'_i \log_2 \frac{x'_i}{y'_i} + \sum_{i=1}^{n} y'_i \log_2 \frac{y'_i}{x'_i}, \\ x'_i &= \frac{|x_i|}{\sum\limits_{i=1}^{n} |x_i|}, y'_i = \frac{|y_i|}{\sum\limits_{i=1}^{n} |y_i|} \end{aligned} \tag{15}$$

(11) Earth mover's distance (EMD)

The EMD distance, which can be understood as the minimal cost for transforming one feature distribution into another [3], is defined below:

Find a set of f_{ij} that minimizes the overall cost:

$$d(X,Y) = \min \sum_i \sum_j d_{ij} f_{ij}, \tag{16}$$

subject to the following constraints:

$$\begin{aligned} & f_{ij} \ge 0, x_i \ge 0, y_j \ge 0, \\ & \sum_i f_{ij} \le y_j, \\ & \sum_j f_{ij} \le x_i, \\ & \sum_i \sum_j f_{ij} = \min \left(\sum_i x_i, \sum_j y_j \right). \end{aligned}$$

where $d_{ij} = 1 - corr(w_i, w_j)$, and $corr(w_i, w_j)$ means the correlation coefficient between the NMF basis vectors w_i and w_j.

This metric has been used in problems where models were non-negative feature distributions, such as color histograms.

(12) Mahalanobis angle distance

$$d(X,Y) = 1 - \frac{X'\sigma^{-1}Y}{\sqrt{X'\sigma^{-1}X}\sqrt{Y'\sigma^{-1}Y}} \tag{17}$$

(13) Chi square distance

$$d(X,Y) = \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{x_i + y_i} \tag{18}$$

(14) Exponential similarity coefficient-based distance

$$d(X,Y) = 1 - \gamma^2(X,Y), \;\; \gamma(X,Y) = \frac{1}{n} \sum_{i=1}^{n} e^{-\frac{3}{4} \frac{(x_i - y_i)^2}{s_i^2}} \tag{19}$$

(15) Non-parametric similarity coefficient-based distance

$$d(X,Y) = 1 - \gamma^2(X,Y), \ \ \gamma(X,Y) = \frac{n_+ - n_-}{n_+ + n_-} \quad (20)$$

here $x_i' = x_i - \overline{x}_i, y_i' = y_i - \overline{y}_i$, n_+ means the frequency of $\{x_i'y_i' \geq 0, \ \ i = 1, \cdots, n\}$, and n_- means the frequency of $\{x_i'y_i' < 0, \ \ i = 1, \cdots, n\}$.

(16) Cosine distance

$$d(X,Y) = 1 - \cos(X,Y) = 1 - \frac{\sum\limits_{i=1}^{n} x_i y_i}{\sqrt{\left(\sum\limits_{i=1}^{n} x_i^2\right)\left(\sum\limits_{i=1}^{n} y_i^2\right)}} \quad (21)$$

(17) Correlation coefficient-based distance (CCBD)

$$d(X,Y) = 1 - \gamma(X,Y),$$
$$\gamma(X,Y) = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{[\sum\limits_{i=1}^{n}(x_i - \overline{x})^2][\sum\limits_{i=1}^{n}(y_i - \overline{y})^2]}} \quad (22)$$

The above four distance measures are all similarity coefficient-based distances. Two distance measures are considered which originated from the theory of multivariate clustering analysis [7] and seem not to have been used in face recognition. Since they are derived from a similarity coefficient specifically defined for non-negative vectors, they will be suitable distance measures for NMF application.

(18) Non-negative vector similarity coefficient-based (N-VSC1) distance 1

$$d(X,Y) = 1 - \gamma^2(X,Y), \ \ \gamma(X,Y) = \frac{\sum\limits_{i=1}^{n} \min(x_i, y_i)}{\sum\limits_{i=1}^{n} \max(x_i, y_i)} \quad (23)$$

(19) Non-negative vector similarity coefficient-based (N-VSC2) distance 2

$$d(X,Y) = 1 - \gamma^2(X,Y), \ \ \gamma(X,Y) = \frac{\sum\limits_{i=1}^{n} \min(x_i, y_i)}{\sum\limits_{i=1}^{n} (x_i + y_i)/2} \quad (24)$$

The Euclidean distance, the Mahalanobis distance and Manhattan distance are the most widely-used in pattern recognition among all the above distance functions.

IV. Testing Databases Used in this Paper

A. CBCL database

The MIT-CBCL face recognition database [15] involves face images of 10 subjects which are classified into two sets: high resolution pictures, and synthetic images (324/subject) rendered from 3D head models of the 10 subjects. In the paper, the second set is used which includes images varying in illumination and pose.

B. CMU AMP face expression database

There are 13 subjects and each one has 75 images showing different expressions in the database (AMP Lab, CMU).

C. YaleB database

The Yale Face Database B (YaleB) [16] consists of 5850 source images of 10 subjects each captured under 585 viewing conditions (9 poses $\times$ 65 illumination conditions). In the preprocess stage, align all frontal pose images by the centers of eyes and mouth and the rest images by the center points of the faces. Then normalize all images with the same resolution 92 $\times$ 112.

D. CBCL2 database

This database [17] is from the Center for Biological and Computational Learning at M.I.T. Similar to the MIT-CBCL face recognition database, it contains gray-scale static images of human face of 10 subjects. There are totally 3080 synthetic images (308/subject) rendered from 3D head models of the 10 subjects.

E. FERET database

The Facial Recognition Technology (FERET) database was sponsored by the Department of Defenses Counter-drug Technology Development Program [18]. We selected 120 persons, 6 frontal-view images for each individual. Face image variations in these 720 images include illumination, facial expression, partial occlusion and aging [18]. All images are aligned by the centers of eyes and mouth and then normalized with resolution 92 $\times$ 112. The pixel value of each image will be normalized to $[0, 1]$.

We use matlab to resize all the images in the above databases to 1/16 of the original size to reduce the computational complexity, then apply NMF algorithm on the downsampled image sets.

V. Experiment

In this section, a face recognition system is built to evaluate the performance of 19 different distance measures using images from databases described in Sect.4. The system applies traditional NMF algorithm for face recognition as in Sect.2. In all the experiments, tr images per person are selected from the database to form a training set and the remainder is the test set. A set of experiments are set on the above system, then the performance of all the distance measures for NMF-based face recognition is assessed.

First, for the NMF algorithm, we compare the two ways to obtain feature vectors (see Sect.II-E.1), then the recognition rates for the five different databases with different experimental settings (tr = 10, 2, 20, 50 and 2; dimensionality of feature vectors at 80, and $proj = 1$ or 2 means the way of feature extraction.) are summarized in Table I. To facilitate comparison, we use bold fonts

TABLE I.
RECOGNITION RATE OF ALL THE DISTANCE MEASURES ($proj = 1, 2$)

Distance measure	CBCL (tr=10, p=80)		CMU AMP (tr=2, p=80)		YaleB (tr=20, p=80)		CBCL 2 (tr=50, p=80)		FERET (tr=2, p=80)	
	$proj = 1$	$proj = 2$	$proj = 1$	$proj = 2$	$proj = 1$	$proj = 2$	$proj = 1$	$proj = 2$	$proj = 1$	$proj = 2$
distance 1	0.8147	0.8879	0.9810	**1.0000**	0.2940	0.2853	0.5806	0.6101	0.6688	0.6292
distance 2	0.8182	0.8920	0.9810	0.9874	0.3066	0.2832	0.6004	0.5934	0.6729	0.6208
distance 3	0.5344	0.6213	0.9579	0.9157	0.2490	0.2527	0.3271	0.3833	0.4688	0.4188
distance 4	0.7917	0.8475	0.9831	0.9947	**0.3696**	0.3561	0.6004	0.6097	0.7042	0.6500
distance 5	0.3850	0.7545	0.8662	0.9737	0.1607	0.2554	0.3841	0.5392	0.5479	0.4875
distance 6	0.3914	0.4395	0.3920	0.6575	0.1046	0.1697	0.4795	0.5450	0.1250	0.1438
distance 7	0.5334	0.8847	0.9515	0.9842	0.2609	0.3618	0.4988	0.5965	0.5771	0.6521
distance 8	0.5430	0.8691	0.9536	0.9842	0.2632	0.3320	0.4992	0.5926	0.5729	0.6667
distance 9	0.8143	0.8873	0.9768	**0.9990**	0.2851	0.3039	0.5419	0.5868	0.6646	0.6479
distance 10	0.8048	0.9127	0.9789	**1.0000**	0.2749	0.3501	0.5415	0.5926	0.6729	0.6875
distance 11	0.1643	0.7656	0.0938	0.9536	0.1062	0.1979	0.1333	0.4810	0.0063	0.3896
distance 12	0.8357	0.9038	**0.9926**	0.9895	0.2131	0.2761	**0.6287**	**0.6194**	**0.7396**	**0.7188**
distance 13	**0.8726**	0.9140	0.9800	**0.9990**	0.3331	0.3239	0.5845	0.6174	0.6896	0.6708
distance 14	0.3392	0.6478	0.7724	0.8714	0.1011	0.1710	0.3190	0.5225	0.2125	0.1875
distance 15	0.1000	0.1000	0.0769	0.0769	0.1000	0.1000	0.1000	0.1000	0.0083	0.0083
distance 16	**0.8662**	**0.9392**	**0.9842**	0.9905	**0.3689**	**0.3899**	**0.6078**	0.6054	**0.7125**	0.6875
distance 17	**0.8503**	**0.9226**	**0.9968**	0.9926	**0.3614**	**0.3869**	**0.6093**	**0.6190**	**0.7167**	**0.6896**
distance 18	0.3901	**0.9436**	0.9831	0.9979	0.1039	**0.3722**	0.5752	**0.6198**	0.6625	**0.6938**
distance 19	0.3768	**0.9436**	0.9831	0.9979	0.1012	**0.3722**	0.5752	**0.6198**	0.6583	**0.6938**

for the best 3 measures in each experimental setting. From the last two rows of Table I, we see that the NVSC1 and NVSC2 distance almost always have the same performance, thus in later experiments we shall just display the results from NVSC1.

From Table I, we can see that:

If $proj = 1$, the best distance measures are the Cosine distance (distance 16) and the CCBD distance (distance 17); and if $proj = 2$, the best distance measures are the CCBD distance and the NVSC distance.

Comparing the results between $proj = 1$ vs. $proj = 2$, then clearly, the second feature extraction method is better overall, especially for the CBCL, CMU AMP and YaleB image databases. (however, the advantage of the first feature extraction approach is less computation load and time, so it's the more common in use.) Some typical cases are further shown in Fig.1, where we vary the dimensionality of feature vectors and just use the best distance measures corresponding to different feature extraction method to make the comparison. From this figure, we see the second approach almost always achieves the better result as p varies. (In the following text, therefore, we shall just focus on the result of the second feature extraction method for comparison.)

We find that the commonly used Euclidean distance (distance 2), Mahalanobis distance (distance 4) and Manhattan distance (distance 1) were not particularly effective based on the second feature extraction method. Among these three popular distance measures, Mahalanobis distance (distance 4) performed best but was ranked in the top 3 in just 1 case (when $proj = 2$). Among all the conventional distance measures (distance 1–6, and 12–17), the CCBD distance (distance 17) achieved the best result and was ranked as one of the best 3 measures in 4 cases (when $proj = 2$).

For the distance measures designed for non-negative vectors, the divergence (distance 7), Kullback-Leibler distance (distance 8), as well as their symmetrized versions (distance 9, 10), were not particularly effective; the EMD distance, which involves linear optimization and therefore takes much more computational time, also failed to obtain a satisfactory result. We obtained the best result by the N-VSC distance (distance 18-19) so far. The NVSC distance was one of the best 3 measures in all (when $proj = 2$) but one case [CMU AMP database, with dimensionality set at 80 and 2 training images]. And even in that case, it was the 4th ranked with a recognition rate 0.9979 ! The NVSC distance was in fact ranked the top performer in 2 cases out of the 5 sets of experiments in addition to being a consistently good performer. Moreover, it's also computationally very efficient since its definition is very simple.

Secondly, for further comprehensive comparison of performance corresponding to different value of tr, using the second approach for feature extraction ($proj = 2$), the dimensionality of the feature vectors is fixed and the recognition rates vs. the value of tr for the CBCL and CMU AMP databases are plotted in Fig.2, where p is the dimensionality of the feature space, and we just concentrate on Euclidean distance, the Manhattan distance, CCBD distance, Mahalanobis distance and the NVSC distance.

In Fig.2, we find that our NVSC distance consistently emerges as the best distance measure across a wide range of tr. So we consider it's better than the CCBD distance corresponding to different experimental setting in face recognition system.

Thirdly, to make another further comparison, we now fix the value of tr and vary p, then obtain the recognition rates for different databases. Some typical cases are shown in Fig.3, where we plot the respective recognition rates (given $proj = 2$) vs. the dimensionality of feature vectors for the CBCL database ($tr = 10$), YaleB database ($tr = 20$), and FERET database ($tr = 2$).

From Fig.3, we see that although the CCBD distance and our NVSC distance achieved similar performance

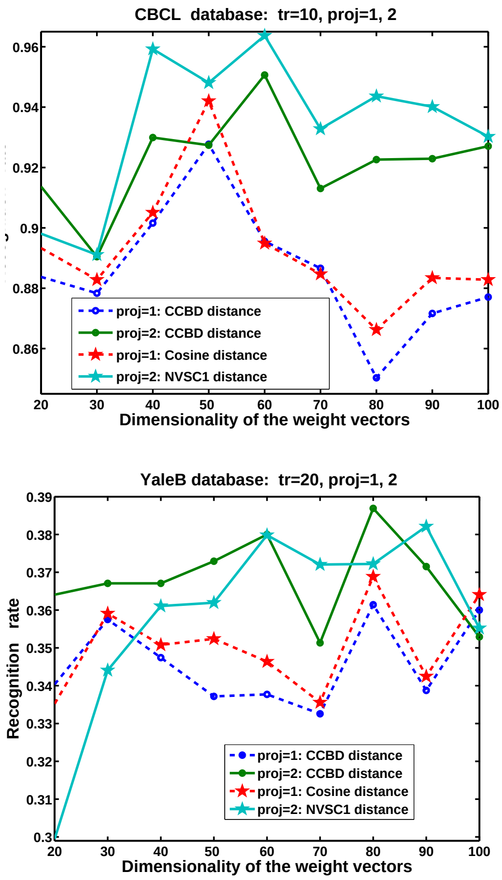

Figure 1. Recognition rate of different feature extraction methods.

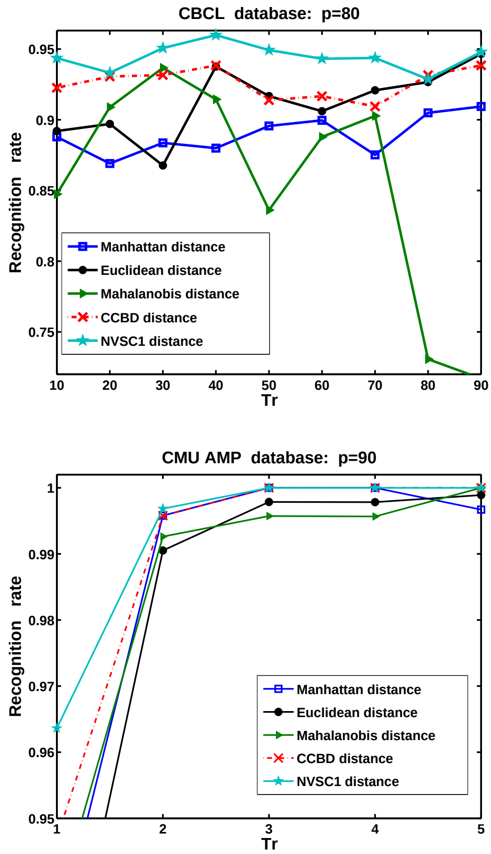

Figure 2. Recognition rate of different distance measures when fixing the dimensionality p.

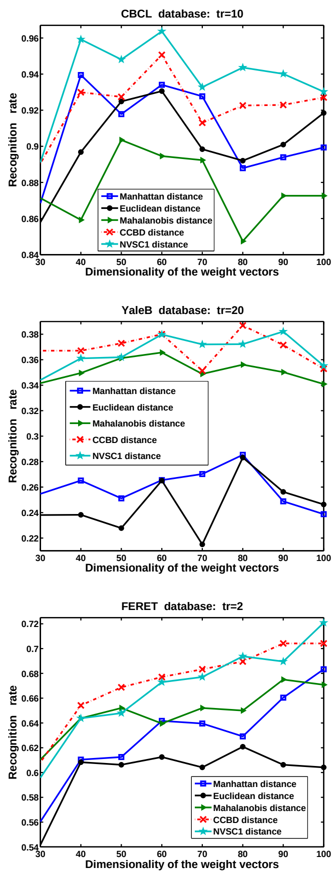

Figure 3. Recognition rate of different distance measures when varying the dimensionality p.

when fixing the value of tr, they both consistently performed better than the Manhattan distance, Euclidean distance, and Mahalanobis distance clearly across a wide range of dimensionality.

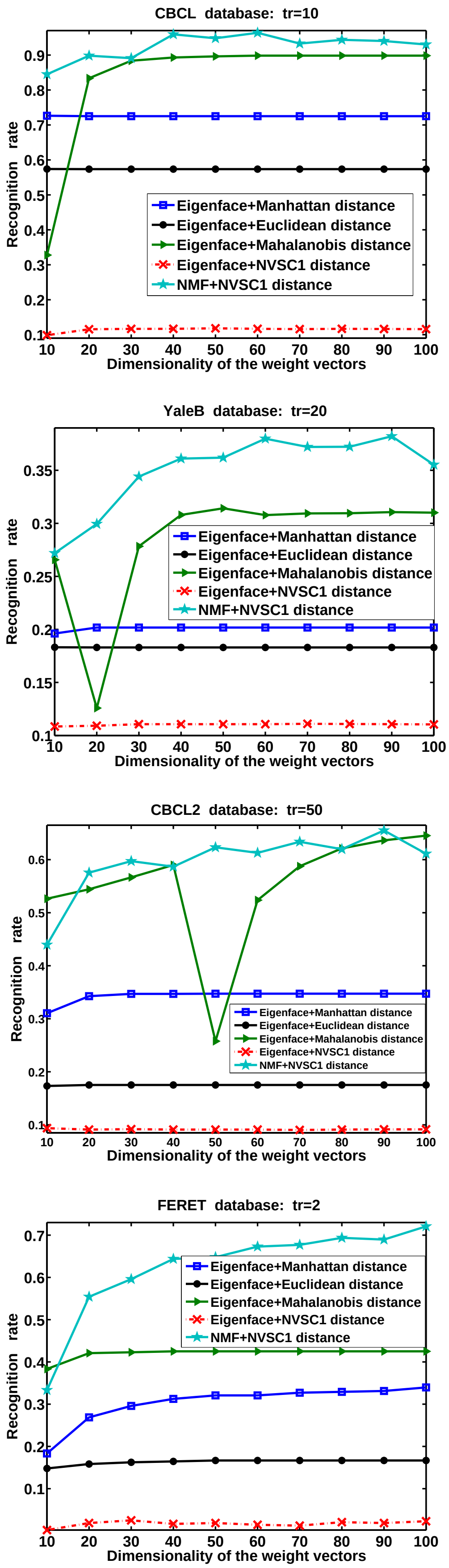

Figure 4. Result comparison with Eigenface.

Finally, to get an overall performance evaluation of our method, we use the Eigenface method as benchmark algorithm. To facilitate the comparison, we just use the Manhattan distance, Euclidean distance, Mahalanobis distance (those are all distance measures in common use, and especially from [11] we know the Mahalanobis distance generally achieve the best recognition results when combined with the Eigenface method) and NVSC1 distance for feature vectors based on the Eigenface method.

The results at different dimensionality and databases are plotted in Fig.4 (the result of CMU AMP database is omitted since the recognition rates of different methods are all too close to 1).

From this figure, we see that:

I. The NVSC1 distance did not lead to a good performance when combined with Eigenface method. In accordance with its definition, it should only be used for non-negative feature vectors and thus is not suitable for use with Eigenface.

II. On all the databases, our NVSC1 distance in conjunction with NMF algorithm ($proj = 2$) always achieves the best result and better than any distance measures combined with Eigenface method.

Based on all the experimental results for face recognition, we conclude that the second feature extraction approach and the NVSC distance are the most suitable for the NMF-based face recognition. Using them, the NMF method performs better than the Eigenface method.

VI. Conclusions & Future Work

As a relative new technique for feature extraction, NMF lacks of a suitable metric distance to work with its non-negative feature vectors. Some traditional distances such as L_1, L_2 and Cosine distance are all in common use for pattern recognition problem, but they do not take into account the positive property of NMF-based feature vector.

In this paper, we compared 19 distance measures for NMF-based face recognition, then showed that it is possible to define a metric for NMF that can remarkably improve the recognition results using the same training set of face images. All the experiments are performed using 5 different face databases. Based on all the experimental results, we concluded:

1) The second NMF-based feature extraction method generally performs better than the first method.
2) Our NVSC distance measure (combined with the second NMF-based feature extraction method) is consistently among the best measures in the face recognition experiments and always performs better than the Euclidean distance, the Mahalanobis distance and the Manhattan distance, which are often used in pattern recognition systems. The effectiveness of the NVSC measure stems from the fact that it is specifically designed for non-negative vectors, so it is the most appropriate for NMF-based face recognition. The entropy-based measures (distance

7-10) can also deal with non-negative vectors, but they are primarily designed for probability distributions and are not effective in coping with vectors with many zero coefficients.

ACKNOWLEDGMENTS

The work is supported by Guangdong Science and Technology Department under Grant No.[2010]170, No.2009B010800051, Science and Technology Bureau of Guangzhou under Grant No. 2011J4300046. The authors would like to thank Prof. J.Y. Yuan (Federal University of Parana, Brazil). Credit is hereby given to the Massachusetts Institute of Technology and to the Center for Biological and Computational Learning for providing the CBCL database of facial images. And thanks to Bernd Heisele at the Center of Biological and Computational Learning at M.I.T. for providing the CBCL2 face database.

REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. Vol. 3, pp. 71–86, 1991.

[2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 401:788–791, 1999.

[3] D. Guillamet and J.Vitrià, "Evaluation of distance metrics for recognition based on non-negative matrix factorization," *Pattern Recogn. Lett.*, vol. 24, no. 9-10, pp. 1599–1605, 2003.

[4] ——, "Determining a suitable metric when using non-negative matrix factorization." in *ICPR (2)*, 2002, pp. 128–131.

[5] Y. Xue, C. S. Tong, and W. Zhang, "Evaluation of distance measures for nmf-based face recognition," in *Proceedings of the 2006 International Conference on Computational Intelligence and Security–Part I*, 2006, pp. 651–656.

[6] ——, "Survey of distance measures for nmf-based face recognition." ser. Lecture Notes in Computer Science. Springer, pp. 1039–1049.

[7] Y. Zhang and K. Fang, *An Introduction to Multivariate Analysis*. Science Press, Beijing, 1982.

[8] T. Feng, S. Z. Li, H.-Y. Shum, and H. Zhang, "Local non-negative matrix factorization as a visual representation," in *ICDL '02: Proceedings of the 2nd International Conference on Development and Learning*. Washington, DC, USA: IEEE Computer Society, 2002, p. 178.

[9] D. Guillamet and J.Vitrià, "Non-negative matrix factorization for face recognition," in *CCIA '02: Proceedings of the 5th Catalonian Conference on AI*. London, UK: Springer-Verlag, 2002, pp. 336–344.

[10] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Adv. Neural Info. Proc. Syst.*, vol. 13, pp. 556–562, 2001.

[11] V. Perlibakas, "Distance measures for PCA-based face recognition," *Pattern Recogn. Lett.*, vol. 25, no. 6, pp. 711–724, 2004.

[12] J. Yu, "Clustering methods, applications of multivariate statistical analysis," School of Electronics Engineering and Computer Science, Peking University, Beijing, 100871, Tech. Rep.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[14] A. Fraser, N. Hengartner, K. Vixie, and B. Wohlberg, "Incorporating invariants in Mahalanobis distance based classifiers: Application to face recognition," in *International Joint Conference on Neural Networks (IJCNN)*, Portland, OR, USA, 2003.

[15] B. Weyrauch, J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3d morphable models," in *The First IEEE Workshop on Face Processing in Video (FPIV 2004)*. Washington, D.C, USA: IEEE, 2004, pp. 1–5.

[16] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Generative models for recognition under variable pose and illumination," in *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2000, pp. 277–284.

[17] J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3d morphable models," in *International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA-03)*, ser. Lecture Notes in Computer Science, J. Kittler and M. S. Nixon, Eds., vol. 2688, no. -. Surrey, UK: Springer, 2003, pp. 27–34.

[18] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000. [Online]. Available: citeseer.ist.psu.edu/article/phillips99feret.html

Yun Xue received his Ph.D. degree in image processing from Hong Kong Baptist University, Hong Kong, in 2007. He is currently an associate professor in South China Normal University and a member of the IEEE and ACM. He has published more than twenty papers. His current research interests include pattern recognition and data mining.