Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Pattern Recognition Letters 25 (2004) 711–724
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Distance measures for PCA-based face recognition

Vytautas Perlibakas *

Image Processing and Multimedia Laboratory, Kaunas University of Technology, Studentu st. 56-305, LT-3031 Kaunas, Lithuania

Received 2 September 2003; received in revised form 9 December 2003

Abstract

In this article we compare 14 distance measures and their modifications between feature vectors with respect to the recognition performance of the principal component analysis (PCA)-based face recognition method and propose modified sum square error (SSE)-based distance. Recognition experiments were performed using the database containing photographies of 423 persons. The experiments showed, that the proposed distance measure was among the first three best measures with respect to different characteristics of the biometric systems. The best recognition results were achieved using the following distance measures: simplified Mahalanobis, weighted angle-based distance, proposed modified SSE-based distance, angle-based distance between whitened feature vectors. Using modified SSE-based distance we need to extract less images in order to achieve 100% cumulative recognition than using any other tested distance measure. We also showed that using the algorithmic combination of distance measures we can achieve better recognition results than using the distances separately.

Keywords: Face recognition; PCA; Distance measures

1. Introduction

Principal component analysis (PCA) or Karhunen–Loeve transform (KLT)-based face recognition method was proposed in (Turk and Pentland, 1991). It was studied by computer scientists (Moon and Phillips, 1998; Yilmaz and Gokmen, 2001; Navarrete and Ruiz-del-Solar, 2001, 2002) and psychologists (Abdi et al., 1995; Hancock et al., 1996), used as a baseline method for comparison of face recognition methods (Moghaddam and Pentland, 1998; Phillips et al., 2000) and implemented in commercial applications (Viisage, 2001). Using PCA we find a subset of principal directions (principal components) in a set of training faces. Then we project faces into this principal components space and get feature vectors. Comparison is performed by calculating the distance between these vectors. Usually comparison of face images is performed by calculating the Euclidean distance between these feature vectors. Sometimes the angle-based distance is used. Mathematical formulation of this recognition method is presented in the next section. Although there exist many other distance measures, we were able to find only few attempts to create, compare and use other distance measures

* Tel./fax: +370-37451577.
E-mail address: vperlib@mmlab.ktu.lt (V. Perlibakas).

0167-8655/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2004.01.011

(Navarrete and Ruiz-del-Solar, 2002; Phillips et al., 1997, 2000) in order to achieve better recognition results.

In this article we compare recognition performance of 14 distance measures including Euclidean, angle-based, Mahalanobis and their modifications. Also we propose modified sum square error-based distance and modified Manhattan distance measures. The experiments showed, that the proposed distance measures were among the best distance measures with respect to different characteristics of the biometric systems. For comparison we used the following characteristics of the biometric systems: equal error rate (EER), first one recognition rate, area above cumulative match characteristic (CMC), area below receiver operating characteristic (ROC), percent of images that we need to extract in order to achieve 100% cumulative recognition.

2. PCA-based face recognition

In this section we will describe Karhunen–Loeve transform (KLT)-based face recognition method, that is often called principal component analysis (PCA) or eigenfaces. We will present only main formulas of this method, whose details could be found in (Groß, 1994).

Let $\boldsymbol{X}_j$ be N-element one-dimensional image and suppose that we have r such images ($j = 1, \dots, r$). A one-dimensional image-column X from the two-dimensional image (face photography) is formed by scanning all the elements of the two-dimensional image row by row and writing them to the column-vector. Then the mean vector, centered data vectors and covariance matrix are calculated:

$$\boldsymbol{m} = \frac{1}{r}\sum_{j=1}^{r}\boldsymbol{X}_j, \tag{1}$$

$$\boldsymbol{d}_j = \boldsymbol{X}_j - \boldsymbol{m}, \tag{2}$$

$$\boldsymbol{C} = \frac{1}{r}\sum_{j=1}^{r}\boldsymbol{d}_j\boldsymbol{d}_j^{\mathrm{T}}, \tag{3}$$

here $\boldsymbol{X} = (x_1, x_2, \dots, x_N)^{\mathrm{T}}$, $\boldsymbol{m} = (m_1, m_2, \dots, m_N)^{\mathrm{T}}$, $\boldsymbol{d} = (d_1, d_2, \dots, d_N)^{\mathrm{T}}$.

In order to perform KLT, it is necessary to find eigenvectors $\boldsymbol{u}_k$ and eigenvalues λ_k of the covariance matrix $\boldsymbol{C}(\boldsymbol{C}\boldsymbol{u}_k = \lambda_k\boldsymbol{u}_k)$. Because the dimensionality (N^2) of the matrix $\boldsymbol{C}$ is large even for a small images, and computation of eigenvectors using traditional methods is complicated, dimensionality of matrix $\boldsymbol{C}$ is reduced using the decomposition described in (Kirby and Sirovich, 1990). Found eigenvectors $\boldsymbol{u} = (u_1, u_2, \dots, u_N)^{\mathrm{T}}$ are normed and sorted in decreasing order according to the corresponding eigenvalues. Then these vectors are transposed and arranged to form the row-vectors of the transformation matrix $\boldsymbol{T}$. Now any data $\boldsymbol{X}$ can be projected into the eigenspace using the following formula:

$$\boldsymbol{Y} = \boldsymbol{T}(\boldsymbol{X} - \boldsymbol{m}), \tag{4}$$

here $\boldsymbol{X} = (x_1, x_2, \dots, x_N)^{\mathrm{T}}$, $\boldsymbol{Y} = (y_1, y_2, \dots, y_r, 0, \dots, 0)^{\mathrm{T}}$.

Also we can perform "whitening" (Bishop, 1995) transform:

$$\boldsymbol{Y} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{T}(\boldsymbol{X} - \boldsymbol{m}), \tag{5}$$

here $\boldsymbol{\Lambda}^{-1/2} = \mathrm{diag}\left(\sqrt{1/\lambda_1}, \sqrt{1/\lambda_2}, \dots, \sqrt{1/\lambda_r}\right)$.

Whitening is a linear rescaling that makes the transformed input data to have zero mean and a covariance matrix given by the identity matrix. For projection into eigenspace we can use not all found eigenvectors, but only a few of them, corresponding to the largest eigenvalues. We can manually select desired number of eigenvectors or use the method described in (Swets et al., 1998).

When the image (human face photography) is projected into the eigenspace we get its eigenfeature vector $\boldsymbol{Z} = (z_1, z_2, \dots, z_n)^{\mathrm{T}} = (y_1, y_2, \dots, y_n)^{\mathrm{T}}$, here n is the number of features. When we have feature vector $\boldsymbol{Z}$ of each face, identification can be performed. After projecting a new unknown face image into the eigenspace we get its feature vector $\boldsymbol{Z}_{\mathrm{new}}$ and calculate the Euclidean distances between unknown face and each known face $\varepsilon_i = \|\boldsymbol{Z}_{\mathrm{new}} - \boldsymbol{Z}_i\|$ and say that the face with projection $\boldsymbol{Z}_{\mathrm{new}}$ belongs to a person $s = \arg\min_i[\varepsilon_i]$. For rejection of unknown faces a threshold τ is chosen and it is said that the face with projection $\boldsymbol{Z}_{\mathrm{new}}$ is unknown if $\varepsilon_s \geqslant \tau$. Distance between projections $\boldsymbol{Z}$ is usually measured using the Euclidean distance, some authors measured the distance between the feature vectors in the eigenspace using the angle-based measure (Phillips et al., 1997), but other distance measures also could be used.

3. Distance measures

Let $\boldsymbol{X}$, $\boldsymbol{Y}$ be eigenfeature vectors of length n. Then we can calculate the following distances between these feature vectors (Grudin, 1997; Yambor and Draper, 2002; Phillips et al., 1999, 2000; Cekanavicius and Murauskas, 2002):

(1) Minkowski distance (L_p metrics)

$$d(\boldsymbol{X},\boldsymbol{Y}) = L_p(\boldsymbol{X},\boldsymbol{Y}) = \left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p}, \tag{6}$$

here $p > 0$;

(2) Manhattan distance (L_1 metrics, city block distance)

$$d(\boldsymbol{X},\boldsymbol{Y}) = L_{p=1}(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{n}|x_i - y_i|; \tag{7}$$

(3) Euclidean distance (L_2 metrics)

$$d(\boldsymbol{X},\boldsymbol{Y}) = L_{p=2}(\boldsymbol{X},\boldsymbol{Y}) = \|\boldsymbol{X}-\boldsymbol{Y}\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}; \tag{8}$$

(4) Squared euclidean distance (sum square error, SSE), mean square error (MSE)

$$d(\boldsymbol{X},\boldsymbol{Y}) = L_{p=2}^2(\boldsymbol{X},\boldsymbol{Y}) = \mathrm{SSE} = \|\boldsymbol{X}-\boldsymbol{Y}\|^2 = \sum_{i=1}^{n}(x_i - y_i)^2; \tag{9}$$

$$d(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{n}L_{p=2}^2(\boldsymbol{X},\boldsymbol{Y}) = \mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2; \tag{10}$$

(5) Angle-based distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = -\cos(\boldsymbol{X},\boldsymbol{Y}), \tag{11}$$

$$\cos(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{n}x_iy_i}{\sqrt{\sum_{i=1}^{m}x_i^2\sum_{i=1}^{m}y_i^2}};$$

(6) Correlation coefficient-based distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = -r(\boldsymbol{X},\boldsymbol{Y}), \tag{12}$$

$$r(\boldsymbol{X},\boldsymbol{Y}) = \frac{n\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{\left(n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right)\left(n\sum_{i=1}^{n}y_i^2 - \left(\sum_{i=1}^{n}y_i\right)^2\right)}};$$

(7) Mahalanobis distance and Mahalanobis distance between normed vectors

$$d(\boldsymbol{X},\boldsymbol{Y}) = -\sum_{i=1}^{n}z_ix_iy_i, \tag{13}$$

$$d(\boldsymbol{X},\boldsymbol{Y}) = -\frac{1}{\sqrt{\sum_{i=1}^{n}x_i^2}\sqrt{\sum_{i=1}^{n}y_i^2}}\sum_{i=1}^{n}z_ix_iy_i, \tag{14}$$

here $z_i = \sqrt{\frac{\lambda_i}{\lambda_i+\alpha^2}}$, $\alpha = 0.25$, λ_i—corresponding eigenvalues, or simplified Mahalanobis distance versions with $z_i = \sqrt{\lambda_i/(\lambda_i+\alpha^2)} \simeq \sqrt{1/\lambda_i}$;

(8) Weighted Manhattan distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{n}z_i|x_i - y_i|, z_i = \sqrt{1/\lambda_i}; \tag{15}$$

(9) Weighted SSE distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{n}z_i(x_i - y_i)^2, \quad z_i = \sqrt{1/\lambda_i}; \tag{16}$$

(10) Weighted angle-based distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = -\frac{\sum_{i=1}^{n}z_ix_iy_i}{\sqrt{\sum_{i=1}^{m}x_i^2\sum_{i=1}^{m}y_i^2}}, \quad z_i = \sqrt{1/\lambda_i}; \tag{17}$$

(11) Chi square distance

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

(12) Canberra distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{n}\frac{|x_i - y_i|}{|x_i| + |y_i|}; \tag{19}$$

(13) Modified Manhattan distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{n}|x_i - y_i|}{\sum_{i=1}^{n}|x_i|\sum_{i=1}^{n}|y_i|}; \tag{20}$$

(14) Modified SSE-based distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i^2}; \tag{21}$$

(15) Weighted modified Manhattan distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{n}z_i|x_i - y_i|}{\sum_{i=1}^{n}|x_i|\sum_{i=1}^{n}|y_i|}, \quad z_i = \sqrt{1/\lambda_i}; \tag{22}$$

(16) Weighted modified SSE-based distance

$$d(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{n}z_i(x_i - y_i)^2}{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i^2}, \quad z_i = \sqrt{1/\lambda_i}; \tag{23}$$

If the feature vectors $\boldsymbol{X}$ are stored in the database, some of the components like $\sum_{i=1}^{n}x_i^2$, $n\sum_{i=1}^{n}x_i^2$, $\left(\sum_{i=1}^{n}x_i\right)^2$, $\sum_{i=1}^{n}|x_i|$ of the described distances could be calculated in advance and stored in the database in order to speed-up comparisons and search in the database. In some cases instead of using eigenvalues λ_i in the distance measures, we can include them in the transformation formula. For example instead of using weighted Manhattan distance (15) we can use whitening transform (5) and simple Manhattan distance (7):

$$\begin{aligned} d_{\text{Weighted Manhattan}}(\boldsymbol{X},\boldsymbol{Y}) &= \sum_{i=1}^{n}\lambda_i^{-1/2}|x_i - y_i| \\ &= \sum_{i=1}^{n}\left|\lambda_i^{-1/2}x_i - \lambda_i^{-1/2}y_i\right| \\ &= \sum_{i=1}^{n}|u_i - v_i|, \end{aligned}$$

here $u_i = \lambda_i^{-1/2}x_i$, $v_i = \lambda_i^{-1/2}y_i$.

That is in some cases instead of using weighted distances we can calculate weighted vectors in advance and then use plain (unweighted) distances. But it must be noted that, although mentioned weighted distances perform some data scaling along principal directions, weighting is not necessarily related with whitening because of different scaling factors. Also it must be noted that some of the distances (e.g. (11) and (12)) could be shifted in order to have positive distance values and scaled in order to have values in the interval $[0, 1]$, but these normalizations increase computation time. So if we do not necessary need normalized values, we can calculate and perform faster search with unnormalized values. When the search is done and we present some of the best results to the user (usually only a small part of the database), then we can normalize the displayed results.

Now we will compare some of the mentioned distance measures using the PCA-based face recognition method.

4. Experiments and results

For experiments we used images from the AR (AR, 1998; Martinez and Benavente, 1998), Bern (1995), BioID (2001), Yale (1997), Manchester (1998), MIT (MIT, 1989; Turk and Pentland, 1991), ORL (ORL, 1992; Samaria and Harter, 1994), Umist (Umist, 1997; Graham and Allinson, 1998), FERET (Phillips et al., 1997) databases. From these databases we collected the database containing photographies of 423 persons (two images per person—one for learning and one for testing). In order to avoid recognition errors related to incorrectly detected faces we manually selected the centers of eyes and lips. Then we rotated the images in order to make the line connecting eye centers horizontal, resized the images and made the distances between the centers of the eyes equal to 26 pixels, calculated the center of the face using the centers of eyes and lips, cropped 64×64 central part of the face, performed histogram equalization on the cropped part of the image. It must be noted, that in some cases histogram equalization reduces recognition performance, but usually it is used in order to normalize illumination. Using the cropped templates we performed PCA-based face recognition. In all the experiments we use the same templates and change only the distance measures between eigenfeature vectors and the number (percent) of used features. For comparison we use cumulative match characteristic (CMC) and receiver operating characteristic (ROC)-based measures, described in (Bromba, 2003).

The results of experiments are summarized in Tables 1–5. In these tables we can see how different distance measures affect recognition accuracy. For measuring overall goodness of the distance measure with respect to recognition accuracy, we use the area above cumulative match characteristic

Table 1
Recognition using 10% of features (42)

Distance measure	Rank (%) of images needed to extract in order to achieve some cumulative recognition percent					CMCA, $0–10^4$	First1 recognition, %	EER, %	ROCA, $0–10^4$
	80	85	90	95	100				
Euclidean; SSE	0.2	0.5	2.6	9.5	85.8	222.16	81.56	7.09	260.84
Angle; Mahalanobis normed	0.2	0.5	1.9	5.9	47.0	117.03	81.56	6.86	139.02_5
Correlation	0.2	0.5	1.4	6.6	53.4	125.02	80.61	6.62	144.80
SSE modified	0.5	0.5	1.2	4.7	24.6_1	95.99	79.91	7.09	162.73
Manhattan	0.2	0.5	2.4	6.4	77.3	128.12	82.27	7.09	218.17
Manhattan modified	0.2	0.5	0.9	2.8	31.4_4	91.07_5	82.27	5.44_5	139.70
Mahalanobis	0.5	0.9	1.7	5.2	46.8	115.60	72.81	6.86	200.56
Mahalanobis simplified	0.2	0.5	0.7	1.9	25.3_2	66.17_1	82.98_5	3.78_3	71.43_4
Angle weighted	0.2	0.5	0.9	2.4	31.0_3	68.32_2	84.63_3	4.49_4	57.65_1
Manhattan weighted	0.2	0.7	2.1	6.9	65.5	149.28	81.09	9.22	307.51
Manhattan weighted modified	0.2	0.5	0.7	2.6	62.4	136.20	81.56	5.91	204.50
SSE weighted	0.2	0.5	1.9	5.4	68.8	123.32	83.22_4	7.09	220.79
SSE weighted modified	0.2	0.5	0.9	2.6	50.4	92.19	82.27	6.15	169.19
Angle whitened	0.2	0.2	0.7	1.9	35.5_5	77.41_4	85.34_1	3.55_1	58.66_2
Correlation whitened	0.2	0.2	0.7	2.1	36.4	77.04_3	85.11_2	3.55_2	59.53_3

Table 2
Recognition using 20% of features (85)

Distance measure	Rank (%) of images needed to extract in order to achieve some cumulative recognition percent					CMCA, $0–10^4$	First1 recognition, %	EER, %	ROCA, $0–10^4$
	80	85	90	95	100				
Euclidean; SSE	0.2	0.5	2.4	7.6	86.1	210.61	83.22	7.33	254.60
Angle; Mahalanobis normed	0.2	0.5	1.4	5.2	49.4_4	106.47	83.45	5.91_5	125.52_5
Correlation	0.2	0.5	1.2	5.4	50.8	109.07	83.22	5.91	127.68
SSE modified	0.2	0.5	1.2	4.3	22.2_1	86.82_4	81.80	6.86	147.15
Manhattan	0.2	0.5	1.7	4.5	73.3	122.20	83.22	7.57	251.09
Manhattan modified	0.2	0.2	0.5	2.4	51.5	92.44	85.82_5	6.15	144.63
Mahalanobis	0.5	0.5	1.2	4.7	51.3	102.55	78.25	6.38	166.16
Mahalanobis simplified	0.2	0.2	0.5	1.2	31.7_2	56.98_1	86.52_4	3.31_2	50.18_2
Angle weighted	0.2	0.2	0.5	1.7	32.9_3	58.71_2	87.00_3	3.78_4	46.44_1
Manhattan weighted	0.2	0.5	1.4	7.3	92.4	200.19	82.98	10.64	442.58
Manhattan weighted modified	0.2	0.2	0.7	3.3	72.8	149.92	85.11	7.09	210.48
SSE weighted	0.2	0.2	1.2	4.3	68.6	130.19	85.11	7.33	270.42
SSE weighted modified	0.2	0.5	0.9	1.9	40.0	97.11	84.16	6.62	216.36
Angle whitened	0.2	0.2	0.5	1.2	51.1_5	85.03_3	88.42_2	3.07_1	68.52_3
Correlation whitened	0.2	0.2	0.5	1.2	51.3	87.60_5	88.89_1	3.31_3	72.70_4

Table 3
Recognition using 30% of features (127)

Distance measure	Rank (%) of images needed to extract in order to achieve some cumulative recognition percent					CMCA, 0–10^4	First1 recognition, %	EER, %	ROCA, 0–10^4
	80	85	90	95	100				
Euclidean; SSE	0.2	0.5	2.1	8.5	86.1	215.20	83.22	7.33	261.63
Angle; Mahalanobis normed	0.2	0.2	1.4	4.3	46.8_4	103.42	85.11	5.91_5	123.06_5
Correlation	0.2	0.5	1.2	4.7	49.2	105.01	84.87	5.91	124.69
SSE modified	0.2	0.5	0.9	3.5	21.3_1	84.11_3	83.45	7.09	144.36
Manhattan	0.2	0.5	1.2	5.0	78.7	155.20	83.92	8.51	318.38
Manhattan modified	0.2	0.2	0.5	2.1	52.2	103.78	87.23_5	6.38	170.44
Mahalanobis	0.2	0.5	0.9	4.5	48.5_5	98.98_5	80.61	5.91	155.97
Mahalanobis simplified	0.2	0.2	0.5	0.9	23.9_2	54.57_1	87.94_4	3.07_1	45.19_2
Angle weighted	0.2	0.2	0.5	1.4	26.0_3	57.15_2	88.42_2	3.31_2	44.08_1
Manhattan weighted	0.2	0.5	4.3	20.1	92.7	316.44	82.98	13.48	628.08
Manhattan weighted modified	0.2	0.5	0.7	7.8	76.4	208.66	84.16	8.51	301.52
SSE weighted	0.2	0.2	1.2	6.6	78.5	166.27	85.34	8.75	343.28
SSE weighted modified	0.2	0.5	1.2	2.6	54.6	118.65	82.98	7.09	261.83
Angle whitened	0.2	0.2	0.5	2.1	49.9	98.92_4	88.18_3	3.78_3	84.84_3
Correlation whitened	0.2	0.2	0.5	2.4	52.0	101.52	88.89_1	3.78_4	87.78_4

Table 4
Recognition using 60% of features (254)

Distance measure	Rank (%) of images needed to extract in order to achieve some cumulative recognition percent					CMCA, 0–10^4	First1 recognition, %	EER, %	ROCA, 0–10^4
	80	85	90	95	100				
Euclidean; SSE	0.2	0.5	2.1	8.0	86.3	216.78	83.45	7.33	265.36
Angle; Mahalanobis normed	0.2	0.2	1.2	4.0	46.6_4	100.46_4	85.58	5.67_5	119.72_3
Correlation	0.2	0.2	1.4	4.0	48.2	101.49_5	85.82	5.91	120.62
SSE modified	0.2	0.5	1.2	3.3	20.8_1	81.12_3	83.22	6.86	140.54
Manhattan	0.2	0.5	1.4	7.3	81.6	218.97	83.92	9.46	388.59
Manhattan modified	0.2	0.2	0.5	2.1	63.8	113.34	86.76_5	6.62	192.74
Mahalanobis	0.2	0.5	0.9	4.3	48.0_5	95.40	81.56	6.15	145.42
Mahalanobis simplified	0.2	0.2	0.5	0.9	21.3_2	49.01_1	88.65_4	2.84_1	36.04_1
Angle weighted	0.2	0.2	0.5	0.9	23.4_3	52.14_2	89.60_3	3.07_2	38.76_2
Manhattan weighted	0.5	2.1	15.1	48.0	98.3	589.51	79.20	17.26	951.26
Manhattan weighted modified	0.5	0.9	5.2	18.0	92.4	298.53	79.43	11.58	196.63
SSE weighted	0.2	0.5	1.4	8.5	83.5	237.58	84.40	10.17	437.54
SSE weighted modified	0.2	0.7	1.7	6.1	58.4	160.40	82.51	8.75	369.03
Angle whitened	0.2	0.2	0.5	1.9	68.8	126.56	89.83_2	4.73_3	119.73_4
Correlation whitened	0.2	0.2	0.5	1.9	68.6	126.00	89.83_1	4.73_4	120.04_5

(CMCA). Smaller CMCA means better overall recognition accuracy. Also we present how many images (in percents) must be extracted from the database in order to achieve some cumulative recognition rate (80–100%). Smaller values mean that we need to extract fewer images in order to achieve some cumulative recognition rate. Last columns of the table are equal error rate (EER) and the area below receiver operating characteristic (ROCA). Smaller values mean better results.

Table 5
Recognition using 90% of features (381)

Distance measure	Rank (%) of images needed to extract in order to achieve some cumulative recognition percent					CMCA, $0–10^4$	First1 recognition, %	EER, %	ROCA, $0–10^4$
	80	85	90	95	100				
Euclidean; SSE	0.2	0.5	2.1	7.6	87.2	217.54	83.69	7.33	266.20
Angle; Mahalanobis normed	0.2	0.2	0.9	3.8	46.1_4	99.42_5	85.34	5.91_3	118.66_3
Correlation	0.2	0.2	1.2	4.0	46.3_5	100.12	85.11	5.91	119.24_4
SSE modified	0.2	0.5	0.9	3.3	20.8_3	79.70_3	83.45	6.86	140.41
Manhattan	0.2	0.5	2.6	12.1	86.1	270.22	83.92	9.46	423.03
Manhattan modified	0.2	0.2	0.5	2.1	77.8	119.80	88.42_3	6.38	182.42
Mahalanobis	0.2	0.5	0.9	4.3	47.0	94.20_4	81.80	5.91	142.75
Mahalanobis simplified	0.2	0.2	0.5	0.9	18.4_2	44.18_1	89.83_2	2.84_1	30.49_1
Angle weighted	0.2	0.2	0.2	0.7	16.3_1	47.78_2	90.07_1	3.07_2	33.82_2
Manhattan weighted	8.7	23.9	49.4	77.1	99.5	1079.20	66.43	18.44	1199.75
Manhattan weighted modified	1.7	5.2	14.9	37.1	94.8	474.77	72.58	15.84	828.84
SSE weighted	0.2	0.5	2.6	16.5	86.8	301.63	83.92	11.35	497.21
SSE weighted modified	0.5	1.4	3.1	14.7	74.9	227.72	78.01	10.64	486.00
Angle whitened	0.2	0.2	0.7	2.8	89.1	168.75	86.52_4	5.20_4	169.87_5
Correlation whitened	0.2	0.2	0.7	3.5	87.9	172.78	86.52_5	5.20_5	172.27

Also we present recognition rate that is achieved if only the first one (most similar) image from the database is extracted. Larger values mean better result. Graphical representation of the used characteristics is shown in Figs. 1 and 2.

Also we performed some experiments using Minkowski distance (6) in order to find out how parameter p influences recognition performance. As we can see from Fig. 3, the best recognition performance is achieved using $p \in [1, 2]$.

Using the results of experiments we can sort the distance measures with respect to the recognition performance using the measured biometric characteristics: overall recognition accuracy (area above cumulative match characteristic—CMCA), first one recognition rate (First1), percent of images needed to extract in order to achieve 100% cumulative recognition (Cum100), equal error rate (EER), area below receiver operating characteristic (ROCA). The results are presented in Table 6. Also the best results are denoted using subscript numbers in Tables 1–5. The relationship between the number of used features and recognition performance is presented in Fig. 4. The best recognition results were achieved using the following distance measures: simplified Mahalanobis, weighted angle-based distance, proposed modified SSE-based distance, angle-based distance between whitened vectors. The proposed modified SSE-based distance measure is among the first three best measures with respect to different characteristics of the biometric systems. Using the proposed modified SSE-based distance we need to extract

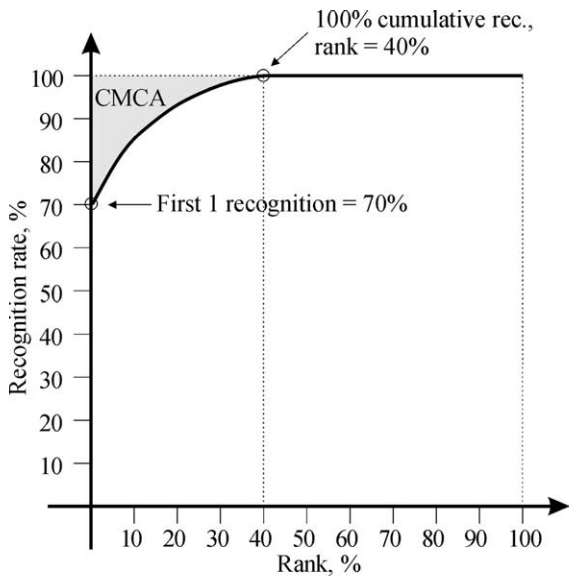

Fig. 1. Cumulative match characteristic.

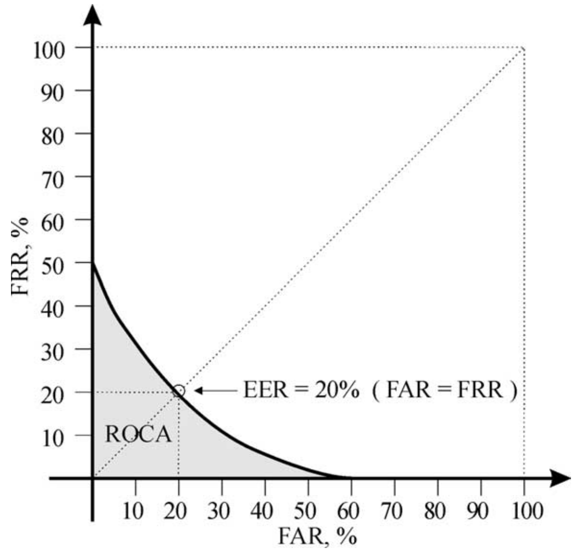

Fig. 2. Receiver operating characteristic.

fewer images (20.8–24.6%) in order to achieve 100% cumulative recognition than using all other tested distance measures if 10–60% of features are used. If we use larger number of features (90%), better 100% cumulative recognition results are achieved using weighted angle (16.3%) or simplified Mahalanobis distance (18.4%). The best results with respect to CMCA are achieved using simplified Mahalanobis distance (44.18–66.17). The largest first one recognition rate is achieved using angle between whitened vectors (85.34%) if we use 10% of features, correlation between whitened vectors (88.89–89.83%) if we use 20–60% of features, weighted angle-based distance (90.07%) if we use 90% of features. The best results with respect to EER are achieved using angle-based distance between whitened vectors (3.07–3.55%) if we use 10–20% of features and simplified Mahalanobis distance (2.84–3.07%) if we use 30–90% of features. The best results with respect to ROCA are achieved using weighted angle-based distance (44.08–57.65%) if we use 10–30% of features and using simplified Mahalanobis distance (30.49–36.04%) if we use 60–90% of features.

Also we tested if the differences between distance measures with respect to Cum100, CMCA, First1, EER, ROCA are statistically significant. We used bootstrap resampling with replacement (Efron and Tibshirani, 1993), analysis of variance (ANOVA) and Tukey's honest significant difference (HSD) post hoc test. We used 30% of features (127), the number of bootstrap samples $N = 2000$, significance level $\alpha = 0.001$. Some descriptive statistics are presented in Table 7. In Table 8 we present Tukey's HSD results. Asterisks indicate significantly different means at an alpha level of $\alpha = 0.001$. In these tables distance IDs are as follows: 1—Mahalanobis (simplified), 2—angle

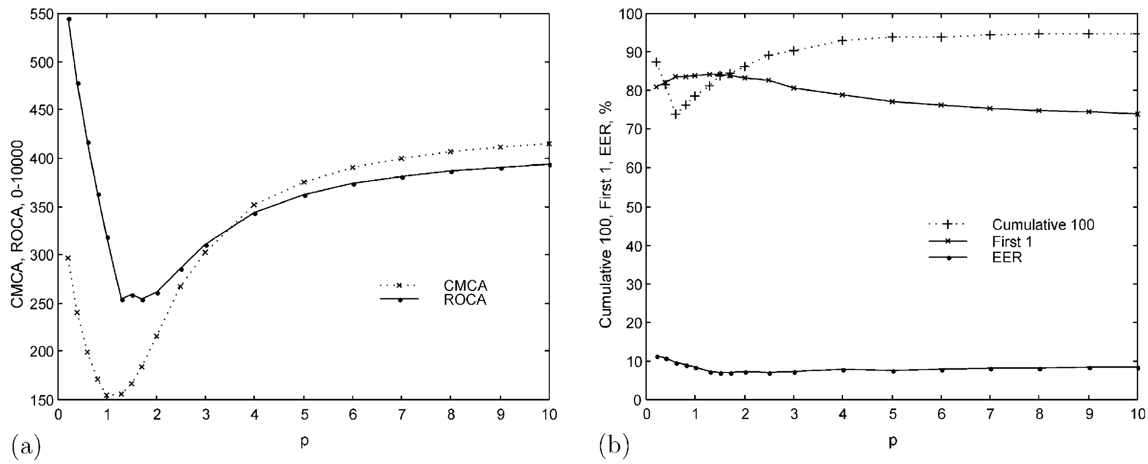

Fig. 3. Recognition performance using Minkowski distance. (a) CMCA and ROCA, (b) Cumulative 100% and First1 recognition, EER.

Table 6
Sorted distance measures with respect to recognition performance

Feat. Num.	CMCA	First1	Cum100	EER	ROCA
10%	Mahalanobis (simplified)	Angle (whitened)	*SSE* (*modified*)	Angle (whitened)	Angle (weighted)
(42)	Angle (weighted)	Correlation (whitened)	Mahalanobis (simplified)	Correlation (whitened)	Angle (whitened)
	Correlation (whitened)	Angle (weighted)	Angle (weighted)	Mahalanobis (simplified)	Correlation (whitened)
20%	Mahalanobis (simplified)	Correlation (whitened)	*SSE* (*modified*)	Angle (whitened)	Angle (weighted)
(85)	Angle (weighted)	Angle (whitened)	Mahalanobis (simplified)	Mahalanobis (simplified)	Mahalanobis (simplified)
	Angle (whitened)	Angle (weighted)	Angle (weighted)	Correlation (whitened)	Angle (whitened)
30%	Mahalanobis (simplified)	Correlation (whitened)	*SSE* (*modified*)	Mahalanobis (simplified)	Angle (weighted)
(127)	Angle (weighted)	Angle (weighted)	Mahalanobis (simplified)	Angle (weighted)	Mahalanobis (simplified)
	SSE (*modified*)	Angle (whitened)	Angle (weighted)	Angle (whitened)	Angle (whitened)
60%	Mahalanobis (simplified)	Correlation (whitened)	*SSE* (*modified*)	Mahalanobis (simplified)	Mahalanobis (simplified)
(254)	Angle (weighted)	Angle (whitened)	Mahalanobis (simplified)	Angle (weighted)	Angle (weighted)
	SSE (*modified*)	Angle (weighted)	Angle (weighted)	Angle (whitened)	Angle
90%	Mahalanobis (simplified)	Angle (weighted)	Angle (weighted)	Mahalanobis (simplified)	Mahalanobis (simplified)
(381)	Angle (weighted)	Mahalanobis (simplified)	Mahalanobis (simplified)	Angle (weighted)	Angle (weighted)
	SSE (*modified*)	*Manhattan* (*modified*)	*SSE* (*modified*)	Angle	Angle

(weighted), 3—SSE (modified), 4—angle (whitened), 5—correlation (whitened). As we can see from the Table 8 the mean differences are not significant at the $\alpha = 0.001$ level between the following distance measures: Mahalanobis (simplified) and SSE (modified) with respect to Cum100, Mahalanobis (simplified) and angle (weighted) with respect to ROCA, angle (whitened) and correlation (whitened) with respect to EER and ROCA, Mahalanobis (simplified) and angle (whitened) with respect to First1, angle (weighted) and angle (whitened) with respect to First1. All other differences are statistically significant. We also used more post hoc tests available in SPSS 12.0 package (SPSS, 2003) and almost all the tests showed the same significant (and insignificant) differences. The only exception was Fisher's least significant difference (LSD) test. This test showed that the difference between Mahalanobis (simplified) and angle (whitened) with respect to First1 is statistically significant.

Now we will compare our results with the results of other researchers. The experiments described in (Phillips et al., 1997, 2000; Navarrete and Ruiz-del-Solar, 2002) showed that recognition performance using PCA-based recognition method with angle-based distance measure is better than using the Euclidean distance, using the Euclidean distance we can achieve larger recognition rates than using Manhattan distance, Mahalanobis distance performs better than other mentioned distances. The experiments with Manhattan, Euclidean, angle-based, Mahalanobis distances and different combinations described in

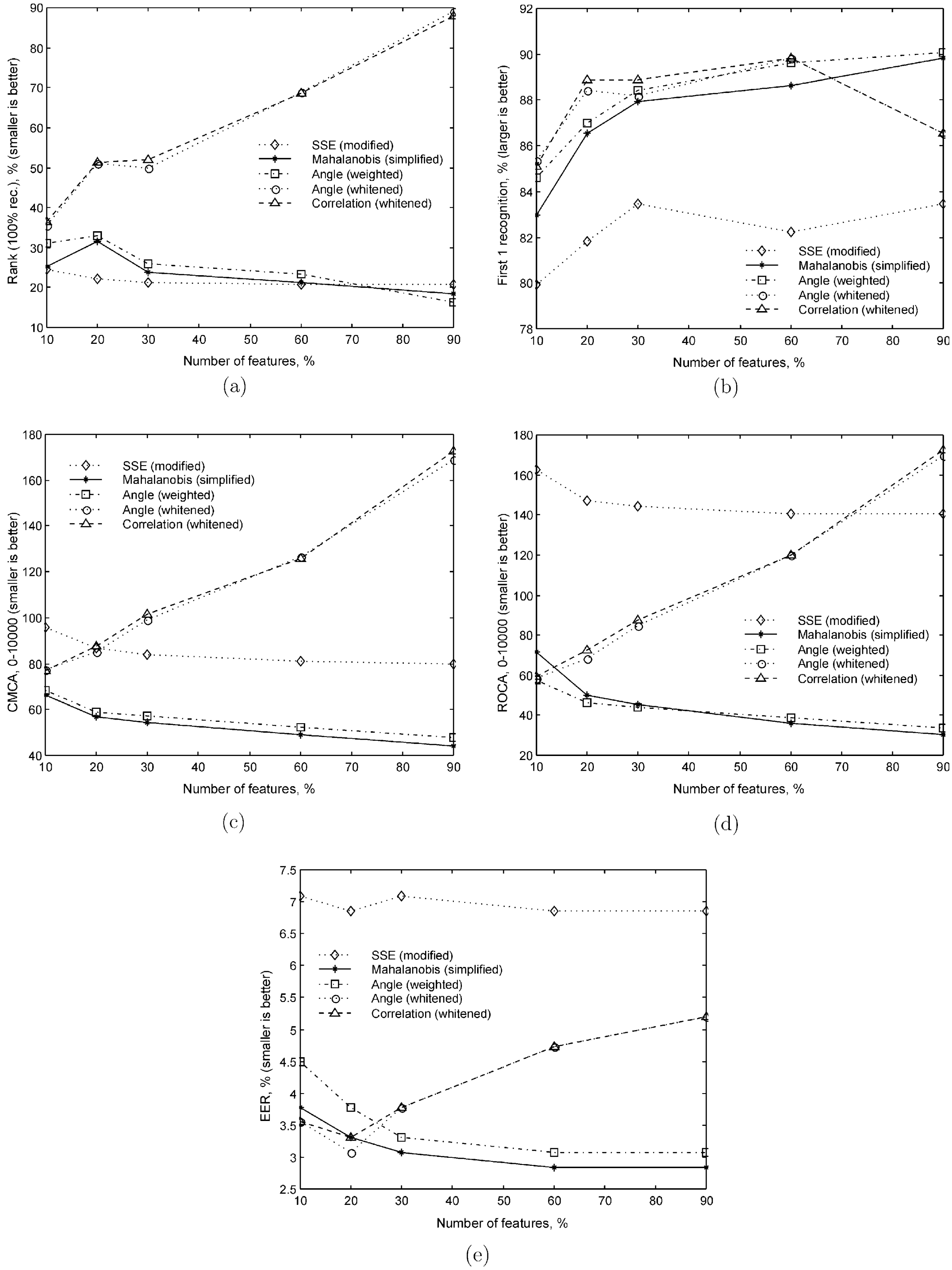

Fig. 4. Recognition performance using different number of features. (a) Cumulative 100% recognition, (b) First1 recognition, (c) CMCA, (d) ROCA and (e) EER.

Table 7
Descriptive statistics

Distance ID	Mean	Std. Dev.	Std. Err.	95% Confidence Interval		Min	Max
				Lower bound	Upper bound		
CMCA							
1	54.4720	9.53966	0.21331	54.0537	54.8904	29.34	90.18
2	56.9940	10.33195	0.23103	56.5409	57.4471	30.04	96.77
3	84.0432	12.61411	0.28206	83.4900	84.5963	48.06	131.64
4	99.0460	21.02596	0.47015	98.1240	99.9680	39.82	174.45
5	101.7862	21.68936	0.48499	100.8350	102.7373	43.76	181.75
First1							
1	89.6742	1.62365	0.03631	89.6030	89.7454	83.69	94.80
2	89.9774	1.59608	0.03569	89.9074	90.0474	84.16	94.56
3	85.1074	1.82273	0.04076	85.0275	85.1874	78.96	91.02
4	89.8616	1.60529	0.03590	89.7912	89.9320	83.92	94.56
5	90.3178	1.55792	0.03484	90.2495	90.3862	84.16	95.04
Cum100							
1	21.6412	3.72947	0.08339	21.4776	21.8047	4.50	30.30
2	24.1841	4.39684	0.09832	23.9913	24.3770	4.50	33.80
3	21.4187	2.00773	0.04489	21.3307	21.5067	11.10	27.90
4	43.6803	8.88014	0.19857	43.2909	44.0698	19.10	60.50
5	45.5153	9.21713	0.20610	45.1111	45.9194	19.10	62.20
EER							
1	3.0896	0.87283	0.01952	3.0513	3.1278	0.97	6.80
2	3.3346	0.76055	0.01701	3.3012	3.3679	1.36	8.78
3	6.7445	1.29171	0.02888	6.6879	6.8012	3.35	15.17
4	3.7646	1.15688	0.02587	3.7139	3.8154	1.05	8.40
5	3.7791	1.18242	0.02644	3.7273	3.8310	1.17	8.85
ROCA							
1	47.0447	19.52074	0.43650	46.1887	47.9007	11.69	170.44
2	44.9181	20.03766	0.44806	44.0394	45.7968	9.32	182.29
3	147.7803	43.57351	0.97433	145.8695	149.6911	56.38	451.57
4	85.7599	36.75803	0.82193	84.1479	87.3718	20.31	279.02
5	88.8438	37.94469	0.84847	87.1798	90.5078	19.53	288.80

(Yambor and Draper, 2002) showed that simplified Mahalanobis distance performs significantly better than L1, L2 or angle-based distance if using more than 60% of eigenfeatures. Our results also showed that angle-based distance performs better than the Euclidean distance. Simplified Mahalanobis distance performs better than the Euclidean, Manhattan and angle-based distance measures with respect to CMCA and EER. But the results also showed, that weighted angle-based distance performs better than simplified Mahalanobis distance with respect to ROCA and first one recognition rate. Also the experiments showed, that the proposed modified SSE-based distance performs better than simplified Mahalanobis distance and weighted angle-based distance with respect to 100% cumulative recognition. We also tested Chi square and Canberra distances, but the results were much worse than using the Euclidean or other tested distance measure. The results using Euclidean or SSE-based distance between whitened feature vectors were worse than the results using angle-based distance between whitened vectors.

In order to achieve larger recognition performance we can try to combine different distance measures as it was done in (Yambor and Draper, 2002). Also we can perform algorithmic

Table 8
Mean differences and significance values

Distance ID 1	Distance ID 2	CMCA	First1	Cum100	EER	ROCA
1	2	−2.52199*	−0.30319*	−2.54300*	−0.24499*	2.12662
		0.000006	0.000000	0.000000	0.000000	0.250291
1	3	−29.57116*	4.56678*	0.22245	−3.65495*	−100.73560*
		0.000000	0.000000	0.801765	0.000000	0.000000
1	4	−44.57398*	−0.18735	−22.03920*	−0.67508*	−38.71515*
		0.000000	0.002906	0.000000	0.000000	0.000000
1	5	−47.31414*	−0.64362*	−23.87410*	−0.68954*	−41.79910*
		0.000000	0.000000	0.000000	0.000000	0.000000
2	3	−27.04916*	4.86998*	2.76545*	−3.40997*	−102.86222*
		0.000000	0.000000	0.000000	0.000000	0.000000
2	4	−42.05199*	0.11584	−19.49620*	−0.43009*	−40.84177*
		0.000000	0.169208	0.000000	0.000000	0.000000
2	5	−44.79215*	−0.34043*	−21.33110*	−0.44455*	−43.92572*
		0.000000	0.000000	0.000000	0.000000	0.000000
3	4	−15.00282*	−4.75414*	−22.26165*	2.97988*	62.02044*
		0.000000	0.000000	0.000000	0.000000	0.000000
3	5	−17.74299*	−5.21040*	−24.09655*	2.96542*	58.93650*
		0.000000	0.000000	0.000000	0.000000	0.000000
4	5	−2.74017*	−0.45626*	−1.83490*	−0.01446	−3.08394
		0.000001	0.000000	0.000000	0.993103	0.026595

Table 9
Recognition using 30% of features (127) and combined distance measures

Distance measure	Rank (%) of images needed to extract in order to achieve some cumulative recognition percent					CMCA, $0–10^4$	First1 recognition, %	EER, %	ROCA, $0–10^4$
	80	85	90	95	100				
SSE (mod.)	0.2	0.5	0.9	3.5	21.3	84.11	83.45	7.09	144.36
Mahalanobis (simplified)	0.2	0.2	0.5	0.9	23.9	54.57	87.94	3.07	45.19
Angle (weighted)	0.2	0.2	0.5	1.4	26.0	57.15	88.42	3.31	44.08
SSE (mod.) + Mahalanobis (simplified)	0.2	0.2	0.5	0.9	16.8	49.54	87.94	3.07	38.97
SSE (mod.) + angle (weighted)	0.2	0.2	0.5	1.4	15.8	50.55	88.42	3.31	36.63
Mahalanobis (simplified) + angle (weighted)	0.2	0.2	0.5	1.4	24.8	56.81	88.42	3.31	41.79

combination (Perlibakas, 2002) by sorting all images using one distance measure (e.g. modified SSE-based distance) and then resorting some part (e.g. 25%) of images with the smallest distances using another distance measure, for example simplified Mahalanobis or weighted angle-based distance. The results of such algorithmic combination are presented in Table 9. As we can see from the table, using such combination we can achieve better performance with respect to CMCA, ROCA and 100% cumulative recognition. But it must be noted that in order to achieve better results using combined method than using not combined methods we must choose an appropriate percent of resorting.

5. Conclusions

In this publication we compared 14 distance measures and their modifications for principal component analysis-based face recognition method and proposed modified sum squared error (SSE)-based distance measure. Recognition experiments were performed using the database containing photographies of 423 persons. The experiments showed, that the proposed distance measure is among the first three best measures with respect to different characteristics of the biometric systems. The best recognition results were achieved using the following distance measures: simplified Mahalanobis, weighted angle-based distance, proposed modified SSE-based distance, angle-based distance between whitened feature vectors. Using the proposed modified SSE-based distance we need to extract less images in order to achieve 100% cumulative recognition than using any other tested distance measure. We also showed that using the algorithmic combination of distance measures we can achieve better recognition results than using the distances separately.

References

Abdi, H., Valentin, D., Edelman, B., O'Toole, A.J., 1995. More about the difference between man and women: Evidence from linear neural network and principal component approach. Perception 24, 539–562.

AR, 1998. The AR face database. Available from <http://rvl1.ecn.purdue.edu/~aleix/ar.html>.

Bern, 1995. Face database at the University of Bern. Available from <ftp://ftp.iam.unibe.ch/pub/Images/FaceImages/>.

BioID, 2001. The BioID face database. Available from <http://www.humanscan.de/support/downloads/facedb.php>.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford. p. 504.

Bromba, M., 2003. Biometrics FAQ. Available from <http://home.t-online.de/home/manfred.bromba/biofaqe.htm>.

Cekanavicius, V., Murauskas, G., 2002. Statistics and its applications. Part 2. TEV, Vilnius. p. 272 (in Lithuanian).

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Graham, D.B., Allinson, G.N.M., 1998. Characterizing virtual eigensignatures for general purpose face recognition. Face recognition: From theory to applications. NATO ASI Series F, Computer and Systems Sciences 163, 446–456.

Groß, M., 1994. Visual computing. The integration of computer graphics, visual perception and imaging. Computer Graphics: Systems and Applications. Springer-Verlag.

Grudin, M.A., 1997. A compact multi-level model for the recognition of facial images. Ph.D. thesis, Liverpool John Moores University.

Hancock, P.J.B., Burton, A.M., Bruce, V., 1996. Race processing: Human perception and principal component analysis. Memory Cognition 24, 26–40.

Kirby, M., Sirovich, L., 1990. Application of the Karhunen–Loeve expansion for the characterization of human faces. IEEE Trans. PAMI 12 (1), 103–108.

Manchester, 1998. Manchester face database. Available from <ftp://peipa.essex.ac.uk/ipa/pix/faces/manchester/>.

Martinez, A.M., Benavente, R., 1998. The AR face database. CVC TR 24.

MIT, 1989. MIT Media Laboratory face database. Available from <ftp://whitechapel.media.mit.edu/pub/images/>.

Moghaddam, B., Pentland, A., 1998. Beyond linear eigenspaces: Bayesian matching for face recognition. Face recognition: From theory to applications. NATO ASI Series F, Computer and Systems Sciences 163, 230–243.

Moon, H., Phillips, P.J., 1998. Analysis of PCA-based face recognition algorithms. In: Bowyer, K.W., Phillips, P.J. (Eds.), Empirical Evaluation Techniques in Computer Vision. IEEE CS Press, pp. 57–71.

Navarrete, P., Ruiz-del-Solar, J., 2001. Eigenspace-based recognition of faces: Comparisons and a new approach. In: International Conference on Image Analysis and Processing ICIAP2001. pp. 42–47.

Navarrete, P., Ruiz-del-Solar, J., 2002. Comparative study between different eigenspace-based approaches for face recognition.

ORL, 1992. The ORL face database at the AT&T (Olivetti) Research Laboratory. Available from <http://www.uk.research.att.com/facedatabase.html>.

Perlibakas, V., 2002. Image and geometry based face recognition. Inform. Technol. Control 1 (22), 73–79.

Phillips, P.J., Moon, H., Rauss, P., Rizvi, S.A., 1997. The FERET September 1996 Database and Evaluation Procedure. In: 1st International Conference on Audio and Video-based Biometric Person Authentication, Crans-Montana, Switzerland.

Phillips, P.J., Moon, H.J., Rizvi, S.A., Rauss, P.J., 2000. The FERET evaluation methodology for face recognition algorithms. PAMI 22 (10), 1090–1104.

Phillips, P.J., O'Toole, A.J., Cheng, Y., Ross, B., Wild, H.A., 1999. Assessing algorithms as computational models for human face recognition. NIST TR.

Samaria, F., Harter, A., 1994. Parameterisation of a stochastic model for human face identification. In: 2nd IEEE Workshop on Applications of Computer Vision.

SPSS, 2003. SPSS Home Page <http://www.spss.com/>.

Swets, D.L., Pathak, Y., Weng, J.J., 1998. An image database system with support for traditional alphanumeric queries and content-based queries by example. Multimedia Tools Appl. (7), 181–212.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. J. Cognit. Neurosci. 3 (1), 71–86.

Umist, 1997. Umist face database. Available from <http://images.ee.umist.ac.uk/danny/database.html>.

Viisage, Inc., 2001. Viisage face recognition technology. Available from <http://www.viisage.com/technology.htm>.

Yale, 1997. The Yale face database. Available from <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.

Yambor, W.S., Draper, B.A., Beveridge, J.R., 2002. Analyzing PCA-based face recognition algorithm: Eigenvector selection and distance measures. In: Christensen, H., Phillips, J. (Eds.), Empirical Evaluation Methods in Computer Vision. World Scientific Press, Singapore.

Yilmaz, A., Gokmen, M., 2001. Eigenhill vs. eigenface and eigenedge. Pattern Recognition 34, 181–184.