Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

An adaptive image Euclidean distance

Jing Li, Bao-Liang Lu*

Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 13 October 2007
Received in revised form 12 June 2008
Accepted 31 July 2008

Keywords:
Image similarity
Image Euclidean distance
Image metric
Gender classification

ABSTRACT

The image Euclidean distance (IMED) considers the spatial relationship between the pixels of different images and can easily be embedded in existing image recognition algorithms that are based on Euclidean distance. IMED uses the prior knowledge that pixels located near one another have little variance in gray scale values, and defines a metric matrix according to the spatial distance between pixels. In this paper, we propose an adaptive image Euclidean distance (AIMED), which considers not only the prior spatial knowledge, but also the prior gray level knowledge from images. The most important advantage of the proposed AIMED over IMED is that AIMED makes the metric matrix adaptive to the content of the concerned images. Two ways of using gray level information are proposed. One is based on gray level distances, and the other is based on cosine dissimilarity of gray levels. Experiments on two facial databases and a handwritten digital database show that AIMED achieves the highest classification accuracy when it is embedded in nearest neighbor classifiers, principal component analysis, and support vector machines.

1. Introduction

Measuring the distance or similarity between images is a fundamental and open problem in both psychology and computer vision. Some psychologists suggest that we human beings judge image similarity in a nonmetric way [1], while some believe that occurs in a manifold way [2]. As in the field of computer vision, the most commonly used distance is Euclidean distance, which converts images into vectors according to gray levels of each pixel, and then compares intensity differences pixel by pixel. Since Euclidean distance discards image structures, it cannot properly represent the real distance between images. If a small variation occurs in similar images, a large Euclidean distance between the images could arise. To overcome this shortcoming of Euclidean distance, various image distances have been proposed in recent years, including histogram cosine distance [3], fractional distance [4], tangent distance [5], Hausdorff distance [6–8], fuzzy feature contrast [9], part-based methods [10,11], Isomap [12], and local linear embedding (LLE) [13]. Among these image distances, Isomap and LLE measure the distance in a manifold way, while others are nonmetric as they do not satisfy all the metric axioms, i.e., self-similarity, symmetry, and the triangle inequality.

On the other hand, applying statistical pattern recognition techniques to computer vision has resulted in significant advancements in recent decades, such as the use of principal component analysis (PCA) and support vector machines (SVMs). Since these techniques are mainly based on the Hilbert space, Euclidean distance has become the most widely used similarity measure. Embedding other image distances into existing pattern classifiers, except for the nearest neighbor classifier, is hard to implement since these image distances are nonmetric. To benefit from the rapid development of pattern recognition techniques, Wang and colleagues [14] have proposed an image Euclidean distance (IMED), which takes the spatial relationship of image pixels into account and is robust to both noise and small deformation [15–20]. Moreover, the task of calculating the IMED of images has been proven to be equivalent to two steps. The first step is to perform a linear transformation on original images, and the second step is to calculate the traditional Euclidean distance between the transformed images. Therefore, IMED can be easily embedded in many existing pattern classifiers such as PCA and SVMs.

IMED uses the prior knowledge that pixels located near one another have little variance in gray levels, and determines the relationship between pixels only according to the distance between pixels on the image lattice. In many applications, however, we are only interested in images in some categories, such as face images or handwritten digital images. Therefore, more prior knowledge can be obtained from these images to determine the relationship between pixels. In this paper, we propose an adaptive image Euclidean distance (AIMED), which makes the image metric adaptive to the content of images by considering both the spatial relationship and the gray level relationship between pixels.

The remainder of the paper is organized as follows. In Section 2, the IMED and how to embed IMED into pattern recognition

* Corresponding author.
E-mail addresses: jinglee@sjtu.edu.cn (J. Li), bllu@sjtu.edu.cn, blu@cs.sjtu.edu.cn (B.-L. Lu).

0031-3203/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.07.017

algorithms are briefly introduced, and the relationship between IMED and manifold distances is also discussed. In Section 3, an AIMED is proposed, and two ways to combine the gray level information are introduced. Section 4 presents the experimental results based on face images and handwritten digit images. Conclusions are given in Section 5.

2. Image Euclidean distance

2.1. Definition

An image with fixed size $M \times N$ can be written as a vector $x=\{x^1,x^2,\dots x^{MN}\}$ according to gray levels of each pixel. The traditional Euclidean distance $d_E(x_1,x_2)$ between vectorized images x_1 and x_2 is defined as

$$d_E^2(x_1,x_2)=\sum_{k=1}^{MN}(x_1^k-x_2^k)^2=(x_1-x_2)^{\mathrm{T}}(x_1-x_2) \tag{1}$$

For traditional Euclidean distance, the assumption that different dimensions x^i and x^j are perpendicular is made, and the relationship between pixels is discarded. As a result, Euclidean distance cannot reflect the real distance between images. On the other hand, the IMED takes the angles between different dimensions into account by introducing the metric matrix G. The IMED $d_{IE}^2(x_1,x_2)$ between images x_1 and x_2 is defined as

$$d_{IE}^2(x_1,x_2)=\sum_{i=1}^{MN}\sum_{j=1}^{MN}g_{ij}(x_1^i-x_2^i)(x_1^j-x_2^j)=(x_1-x_2)^{\mathrm{T}}G(x_1-x_2) \tag{2}$$

where the symmetric and positive definite matrix G is referred to as metric matrix, and g_{ij} is the metric coefficient indicating the spatial relationship between pixels p_i and p_j. The definition of g_{ij} is given by

$$g_{ij}=f(d_{ij}^s)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{(d_{ij}^s)^2}{2\sigma^2}\right) \tag{3}$$

where d_{ij}^s is the *spatial* distance between p_i and p_j on the image lattice, and σ is the width parameter. For example, if p_i is at location (k,l) and p_j is at location (k',l'), then d_{ij}^s is

$$d_{ij}^s=\sqrt{(k-k')^2+(l-l')^2} \tag{4}$$

Since IMED considers the spatial relationship between pixels, it is relatively insensitive to small spatial deformation. For example, two images as shown in Figs. 1(a) and (b) are different, but the image in Fig. 1(c) is slightly deformed from that in Fig. 1(a). Computing the Euclidean distances between these images, we have $d_E(a,b)=5.63$ and $d_E(a,c)=6.31$. While computing IMED, we have $d_{IE}(a,b)=3.91$ and $d_{IE}(a,c)=2.15$. By examining Figs. 1(a)–(c), we can see that IMEDs are more reasonable than Euclidean distances are, because $d_{IE}(a,c)$ is less than $d_{IE}(a,b)$, while $d_E(a,c)$ is larger than $d_E(a,b)$.

2.2. Standardizing transformation

In comparison to other image distances, one prominent characteristic of IMED is that it can be easily embedded into almost all of the existing pattern classifiers. By applying the standardizing transformation $G^{1/2}$ [14] to the original images x_1 and x_2, we have

$$u_1=G^{1/2}x_1 \quad \text{and} \quad u_2=G^{1/2}x_2 \tag{5}$$

Thus, the task of calculating IMED between images x_1 and x_2 can be converted to the calculation of the traditional Euclidean distance between u_1 and u_2 as follows:

$$\begin{aligned}d_{IE}^2(x_1,x_2)&=(x_1-x_2)^{\mathrm{T}}G(x_1-x_2)\\&=(x_1-x_2)^{\mathrm{T}}G^{1/2}G^{1/2}(x_1-x_2)\\&=(u_1-u_2)^{\mathrm{T}}(u_1-u_2)\end{aligned} \tag{6}$$

Following Eq. (6), embedding IMED in a classifier is to simply perform the standardizing transformation to images before feeding them to the classification algorithm. Moreover, most elements in $G^{1/2}$ are nearly zero, and the transformation of $G^{1/2}$ can be viewed as

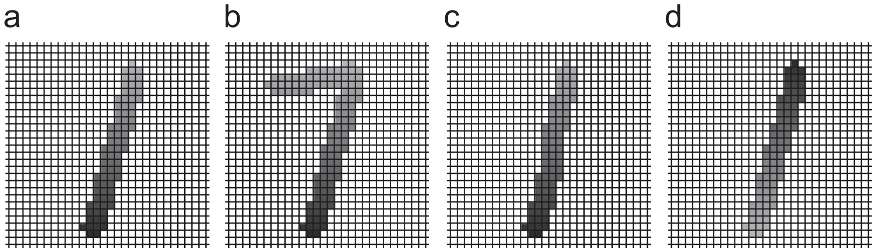

Fig. 1. Similar and dissimilar images. Here, the gray levels of images (a)–(c) become darker from upper row to lower row, while the gray levels of image (d) change in the opposite direction.

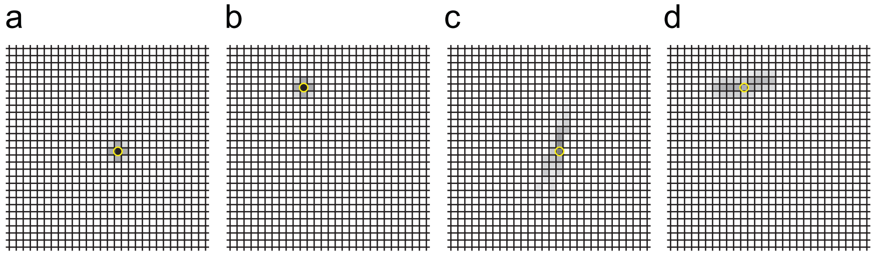

Fig. 2. Masks formed by different image metrics at different locations: (a) IMED, (b) IMED, (c) AIMED-D, and (d) AIMED-D. Here the location of each mask is indicated by a small circle.

a two-dimensional convolution with a small mask. Figs. 2(a) and (b) show the masks at different locations for the digital images shown in Fig. 1. Since g_{ij} is only correlated with the spatial distance d_{ij}^s, the masks are the same at different locations, and the values in the mask are invariant with directions.

2.3. Relationship with manifold ways of image distance

In this subsection, we provide further insights into IMED from the Riemannian geometry [21,22] perspective and discuss the relationship between IMED and manifold distances. Seung and Lee [2] have proposed that images can be considered to be low dimensional manifolds in a high dimensional image space. To measure the distance between two vectors in a manifold, Riemannian geometry is normally used. According to Riemannian geometry, the distance between vector w and vector $w+\mathrm{d}w$ in an M dimension space is defined as

$$d_R^2(w,w+\mathrm{d}w)=\sum_{i=1}^{M}\sum_{j=1}^{M}\mathrm{d}w_i\mathrm{d}w_j g_{ij}(w)=\mathrm{d}w^{\mathrm{T}}G(w)\mathrm{d}w \tag{7}$$

where $G(w)$ is a positive definite matrix and is called the Riemann metric tensor.

For the data distribution of real world images, $G(w)$ changes with w, and can be estimated by the nearest neighbors around w, such as the implementation of LLE [13]. But estimating $G(w)$ is time consuming, and also demands that the distribution of training samples should be dense enough. Another open problem with unfixed $G(w)$ is that the corresponding distance can only be embedded in a nearest neighbor classifier in most cases, and it is hard to be applied to other pattern classifiers such as SVMs.

On the other hand, Eq. (7) can be reduced to Eq. (2), providing that $G(w)$ is independent of w. Thus, the task of calculating the distance between two vectors in Riemann space can be carried out by using Eq. (6), which is rather straightforward. Since $G(w)$ is vector invariant, IMED may not reflect the real data distribution in the image space. IMED can thus be viewed as a tradeoff between precision and computation.

3. Adaptive image Euclidean distance

IMED uses the prior knowledge that pixels located near one another have little variance in gray scale values, and defines the metric matrix G according to the spatial distance between pixels. However, in many applications, only those images from certain categories are of interest, such as the facial images in face recognition problems. These images comprise several objects. For example, a face image often includes eyes, nose, mouth, and so on. The pixels located on the same object may have a closer relationship than that of pixels located on different objects, even though the former have larger spatial distances than do the latter. Therefore, more prior knowledge should be used to define the metric matrix G to adapt to the content of images. From the view of Riemannian geometry, the images concerned constitute a specific manifold in image space. To express the manifold more precisely and to maintain the easily embedded ability of IMED, different vector invariant metric matrices should be used according to different manifolds. In this paper, we propose an AIMED, which is adaptable to the content of the images of interest by considering the relationship between pixels according to both the spatial information and the gray level information. Two methods for combining the gray level information are considered in this paper. One is based on the distance of gray levels between pixels (AIMED-D), and the other is based on the cosine dissimilarity of gray levels between pixels (AIMED-C).

3.1. AIMED-D

Let T be the given set of images

$$T=\{x_l\}_{l=1}^{L} \tag{8}$$

where $x_l \in R^{MN}$ is the l th image, and L is the total number of images. Let T_S be the set of images after performing the standardizing transformation:

$$T_S=\{u_l\}_{l=1}^{L}=\{G^{1/2}x_l\}_{l=1}^{L} \tag{9}$$

To define the gray level distance between different pixels on the given set T, we vectorize the gray level of each image on the pixel p_i as follows:

$$q_i=[x_1^i,x_2^i,\ldots,x_l^i,\ldots,x_L^i] \tag{10}$$

where x_l^i is the gray level value of the l th image on the i th pixel. Then the *gray* level distance of two pixels p_i and p_j on the set T can be defined as the distance of vectors q_i and q_j as follows:

$$d_{ij}^g=\sqrt{(q_i-q_j)^{\mathrm{T}}(q_i-q_j)}=\sqrt{\sum_{l=1}^{L}(x_l^i-x_l^j)^2} \tag{11}$$

By using Eq. (11), we can calculate the gray level distance between pixels based on the original images. However, this gray level distance is sensitive to noise and small spatial deformation. To overcome this deficiency, we redefine the gray level distance between pixels based on the standardizing transformed image set T_S in Eq. (9) as follows:

$$d_{ij}^g=\frac{\sqrt{\sum_{l=1}^{L}(u_l^i-u_l^j)^2}}{L} \tag{12}$$

where u_l^i is the gray level value on the i th pixel of the l th standardizing transformed image.

Following Eq. (3), we define the matrix coefficient g_{ij}^d which represents the gray level relationship between pixels p_i and p_j as follows:

$$g_{ij}^d=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{\eta(d_{ij}^g)^2}{2\sigma^2}\right) \tag{13}$$

where η is a coefficient that makes the contribution of gray scale distance comparable with that of spatial distance. Since a small distance means a close relationship, we can choose η to make the number of small values in d_{ij}^g comparable with that in d_{ij}^s.

According to the discussions mentioned above, the definition of AIMED-D between images x_1 and x_2 can be expressed as

$$d_D^2(x_1,x_2)=(u_1-u_2)^{\mathrm{T}}G_D(u_1-u_2)=\sum_{i=1}^{MN}\sum_{j=1}^{MN}g_{ij}^d(u_1^i-u_2^i)(u_1^j-u_2^j) \tag{14}$$

where u_1 and u_2 are the images after the standardizing transformation of x_1 and x_2, respectively. Comparing Eq. (14) with Eq. (2), we can see that AIMED-D considers not only the spatial relationship between pixels, but also the gray level relationship between pixels. Considering gray level information makes AIMED-D adaptive to the content of the concerned images.

Aside from considering the relationship between pixels, another prominent characteristic of IMED is that it can be easily embedded in existing pattern recognition algorithms by using the standardizing transformation. In the following, we will show that introducing the gray level information will not influence the embeddability of

AIMED-D. Since d_{ij}^{g} satisfies the metric axioms, matrix G_D is symmetric and positive definite [14], and can be decomposed as

$$G_D = G_D^{1/2} G_D^{1/2} \tag{15}$$

Applying the transformation $G_D^{1/2} G^{1/2}$ to the images x_1 and x_2, we have

$$v_1 = G_D^{1/2} u_1 = G_D^{1/2} G^{1/2} x_1$$

and

$$v_2 = G_D^{1/2} u_2 = G_D^{1/2} G^{1/2} x_2 \tag{16}$$

As a result, the AIMED-D between x_1 and x_2 can be reduced to the following traditional Euclidean distance between v_1 and v_2,

$$\begin{aligned} d_D^2(x_1, x_2) &= (u_1 - u_2)^{\mathrm{T}} G_D (u_1 - u_2) \\ &= (u_1 - u_2)^{\mathrm{T}} G_D^{1/2} G_D^{1/2} (u_1 - u_2) \\ &= (x_1 - x_2)^{\mathrm{T}} G^{1/2} G_D^{1/2} G_D^{1/2} G^{1/2} (x_1 - x_2) \\ &= (v_1 - v_2)^{\mathrm{T}} (v_1 - v_2) \end{aligned} \tag{17}$$

Similar to IMED, we can reshape the values in the transformation $G_D^{1/2} G^{1/2}$ to the original image size. Figs. 2(c) and (d) show the two masks in AIMED-D for the digital images in Fig. 1. From Fig. 2, we can see that in IMED, the masks are the same at different locations, and the values in the masks are invariant with direction. They cannot reflect any information about the content of the images, while in AIMED-D, the masks are adaptive to the shape of digital numbers at different locations. Therefore, the masks in IMED smooth the images with a side effect of introducing the influence of the background, while the masks in AIMED-D smooth the gray level differences between pixels only on the same object. Here, we must point out that since the mask in AIMED-D is position-dependent, the transformation of $G_D^{1/2} G^{1/2}$ cannot be viewed as a two-dimensional convolution. One may wonder that AIMED-D cannot take the advantage of being computable by the fast Fourier transform (FFT). In actuality, since the mask size is very small in IMED, the computation implemented directly can be very fast, and there is no need to use the FFT.

Since AIMED-D considers both spatial and gray level relationships between pixels, it is relatively insensitive to both small deformation and gray level deformation. For example, the images shown in Figs. 1(a) and (d) have the same shape but different changes in gray levels, and the images of Figs. 1(a) and (b) have different shapes but the same change in gray levels. Computing the Euclidean distances yields $d_E(a, b) = 5.63$ and $d_E(a, d) = 9.32$. Computing the IMED yields $d_{IE}(a, b) = 3.91$ and $d_{IE}(a, d) = 6.39$. Both Euclidean and IMED cannot lead to satisfactory results. On the contrary, calculating the AIMED-D yields $d_D(a, b) = 19.27$ and $d_D(a, d) = 12.27$. Obviously, AIMED-D is vastly superior to both Euclidean distance and IMED in expressing similarities between images.

3.2. AIMED-C

In the preceding subsection, we defined the relationship between two pixels as the function based on the combination of spatial distance and gray level distance. In this subsection, we consider the relationship between two pixels based on the cosine dissimilarity of gray levels and combine it with spatial distance.

The cosine dissimilarity of two pixels p_i and p_j on a given image set described in Eq. (8) can be defined as

$$c_{ij} = 1 - \frac{q_i^{\mathrm{T}} q_j}{\|q_i\| \|q_j\|} = 1 - \frac{\sum_{l=1}^{L} (x_l^i x_l^j)}{\sqrt{\sum_{l=1}^{L} (x_l^i)^2} \sqrt{\sum_{l=1}^{L} (x_l^j)^2}} \tag{18}$$

where q_i and q_j are defined by Eq. (10). To eliminate the influence from noise and small spatial deformation, we redefine c_{ij} based on the standardizing transformed images as follows:

$$c_{ij} = 1 - \frac{\sum_{l=1}^{L} (u_l^i u_l^j)}{\sqrt{\sum_{l=1}^{L} (u_l^i)^2} \sqrt{\sum_{l=1}^{L} (u_l^j)^2}} \tag{19}$$

where u_l^i is the gray level value on the i th pixel of the l th standardizing transformed image.

Following Eq. (14), the AIMED-C between images x_1 and x_2 can be expressed as

$$d_C^2(x_1, x_2) = (u_1 - u_2)^{\mathrm{T}} G_C (u_1 - u_2) = \sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij}^c (u_1^i - u_2^i)(u_1^j - u_2^j) \tag{20}$$

where u_1 and u_2 are the images after the standardizing transformation of x_1 and x_2, respectively, and g_{ij}^c is defined as

$$g_{ij}^c = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\eta c_{ij}}{2\sigma^2}\right) \tag{21}$$

For images with positive gray values, G_C is symmetric and positive definite. Therefore, G_C can be decomposed as

$$G_C = G_C^{1/2} G_C^{1/2} \tag{22}$$

Similarly, we can apply the transformation $G_C^{1/2} G^{1/2}$ to images x_1 and x_2, and reduce the AIMED-C between original images to the traditional Euclidean distance between the transformed images.

The AIMED-Cs for the images shown in Fig. 1 are calculated as $d_C(a, b) = 32.29$, $d_C(a, c) = 4.80$, and $d_C(a, d) = 18.50$. Examining the images in Fig. 1, we can see that AIMED-Cs are more reasonable than both IMEDs and Euclidean distances.

Finally, we want to discuss the relationship between IMED, AIMED, and the pre-whitening method. Although they all decorrelate the relationship between pixels, they are completely different. The distance between vectors after pre-whitening can be viewed as a special case of the Mahalanobis distance. Wang and colleagues have revealed that the Mahalanobis distance is even more sensitive to small deformation than the traditional Euclidean distance is, and it has a completely opposite behavior as dose IMED [14].

Moreover, there are more restrictions on pre-whitening. For some applications where the number of vectors is fewer than the dimension of vectors, the covariance matrix is not positive definite, therefore the vectors cannot be pre-whitened. For example, the image size of the images shown in Fig. 1 is 28 × 28, which means the dimension of the corresponding vectors is 784. Since there are only four images, they cannot be pre-whitened. But the calculation of the IMED, AIMED-D, and AIMED-C between these images is not influenced by the number of images.

4. Experiments

In order to evaluate our proposed AIMED, we have conducted experiments on two kinds of images. One is the facial images from the CAS-PEAL-R1 face database [23] and the UMIST face database [24], and the other is the handwritten digital images from the MNIST database [25].

In the experiments, G_D and G_C were calculated according to all the training data in each data set. The parameter σ in Eq. (3) affects the spatial correlation among image pixels [16]. A small value of σ means considering little correlation among pixels, and makes the performance of IMED very close to that of the Euclidean distance, while a large value of σ means considering too much correlation among pixels, which would blur the images. Therefore, σ should be chosen according to cross-validation on the training data set. In our experiments, σ was chosen from the set $[0.1, 0.25, 0.5, 1, 1.5, 2, 2.5]$. The parameter η in Eqs. (13) and (21) should be chosen to make the contributions of considering pixel relationships based on gray level and space comparable, and it was chosen to be from 2^6 to 2^{15} in our experiments based on cross-validation on the training data set.

4.1. Gender classification

The CAS-PEAL-R1 face database [23] currently contains 21,832 images of 1040 individuals (595 males and 445 females) in the 'pose' subdirectory. The training and test data sets are collected according to the same rule in Ref. [20]. For this series, 5460 images of 260 individuals whose ID numbers could be evenly divided by four are used as the test data sets, while the remaining 16,372 images of 780 individuals are used as the training data sets. The images of the individuals whose ID numbers are less than 800 in the training data sets are divided into three groups according to poses: looking left (looking left from 22° to 90°), looking straight ahead (from looking left at 22° to looking right as 22°), and looking right (looking right from 22° to 90°). The images of the remaining individuals in the training data sets form the fourth training data set. Detailed information from each data set is listed in Table 1.

Table 1
Number of images in each data set in the CAS-PEAL-R1 face database

No.	Pose	Male	Female	Training	Test
1	Left	3039	2422	3826	1635
2	Mid	3729	3402	4941	2190
3	Right	3039	2422	3826	1635
4	All	5732	3507	3779	5460

In our experiments, images are scaled according to the eye coordinates and cropped to leave only the face area. No masking template was used because we think that the face outlines are important for gender classification, while this information will be removed when using masking template. The final image resolution is 60 × 48 pixels.

Fig. 3 shows the 7 × 7 masks produced by three different methods at different locations, where the masks of AIMED-D and AIMED-C are calculated from the second training data set, which has the face pose of looking straight ahead. We can see that the masks in IMED are the same at different locations, except for pixels that are located on the image border. These masks are not relevant to the content of the facial images and do not reflect any information about the images. While in AIMED-D and AIMED-C, the masks are different at different locations. Only pixels with near intensities are affected with one another. Each mask reflects the shape around each pixel. For example, the mask on the nose is upright; the mask on the mouth is horizontal, as is the shape of lips; the mask on the cheek is nearly round; the mask on the face outline has the same direction as the edge; and the mask on the right eye has almost a single nonzero value at the center point, since the pixels on the eyes always have varied intensity.

After using different transforms on the images according to different image distances, two pattern recognition methods were used to distinguish genders. One was the nearest neighbor classifier, and the other was SVM with radial basis function, a proven classifier in gender classification [20,26,27]. In our experiments, LIBSVM [28] was used for the implementation of SVM, and the parameters in SVMs were chosen by five-fold cross-validation on the training data sets.

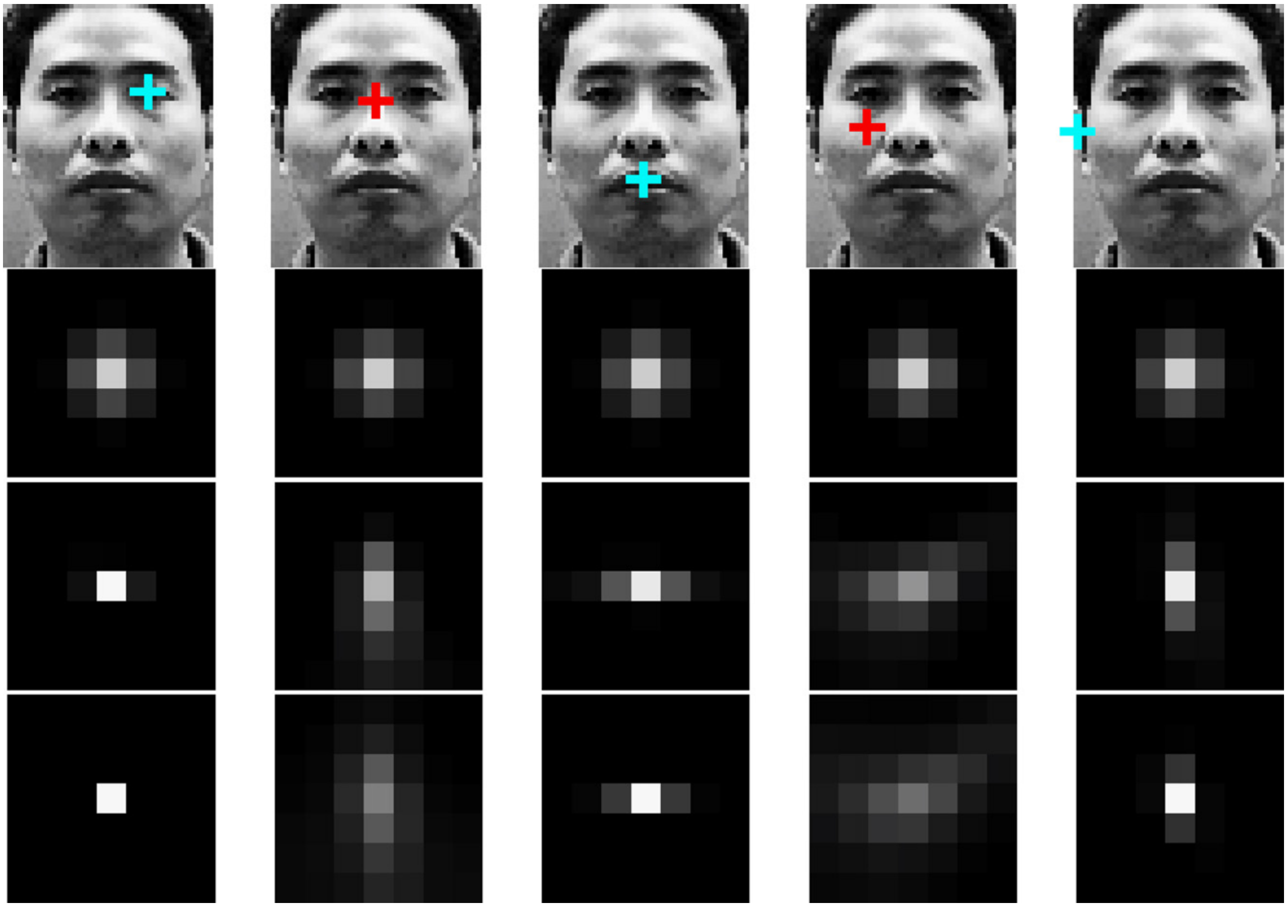

Fig. 3. Masks generated by three different methods at different locations. Here, the first row shows the location of each mask with the mark '+', and the values of $G^{1/2}$, $G_D^{1/2}$, and $G_C^{1/2}$ are shown in second row, third row, and fourth row, respectively.

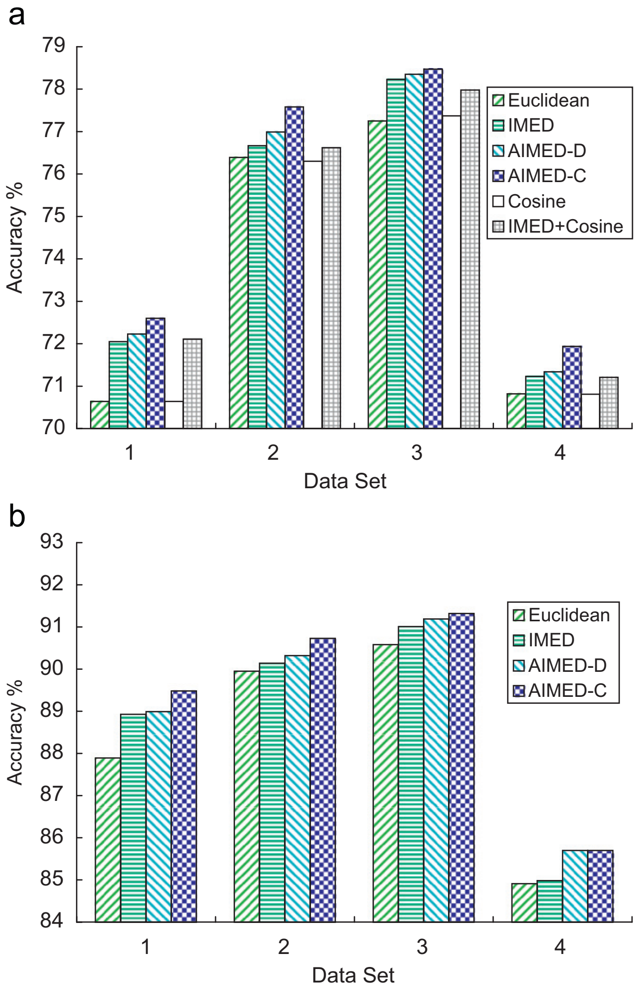

Fig. 4. Classification accuracies with different image metrics: (a) nearest neighbor classifiers and (b) support vector machines.

Table 2
Results of SVMs with different image metrics

No.	Euclidean			IMED			AIMED-D			AIMED-C		
	nSV	Time		nSV	Time		nSV	Time		nSV	Time	
1	2418	284	97	1946	330	79	1230	215	51	**1218**	**163**	**50**
2	3130	688	127	2500	557	102	**1207**	293	**50**	1746	**261**	72
3	2185	253	92	1764	307	73	1145	144	48	**1135**	**140**	**47**
4	2304	385	93	1864	301	78	1452	164	59	**1172**	**143**	**51**

Here, 'nSV' denotes the number of support vectors, the left column in 'Time' denotes the training time (the unit is s), and the right column denotes the test time of each sample (the unit is ms).

The experiments were performed on a 2.8 GHz Pentium 4 PC with 1 GB RAM. The classification results of the two methods are shown in Fig. 4. The number of support vectors and the training time of SVMs are presented in Table 2, where the bold values indicate the least number of support vectors and the shortest training time.

Sine in AIMED-C, the cosine similarity is used to calculate matrix G_C, experiments using cosine similarity between the images and the cosine similarity between their IMED standardized transforms have been performed for comparison. Since the cosine similarity cannot satisfy the triangle inequality, it is nonmetric, and cannot be embedded in SVMs directly. Therefore, we only show the results obtained using nearest neighbor classifiers with the cosine similarity measure.

From Fig. 4, we can see that IMED performs better than traditional Euclidean distance, AIMED-D performs better than IMED, and AIMED-C achieves the best performance whether embedded in nearest neighbor classifiers or SVMs. When nearest neighbor classifiers are used, the average accuracies of AIMED-D and AIMED-C increase by 0.95% and 1.37%, respectively, compared to those of traditional Euclidean distance and by 0.18% and 0.60%, respectively, compared to those of IMED. When SVMs are used, the average accuracies of AIMED-D and AIMED-C increase by 0.72% and 0.98%, respectively, compared to those of traditional Euclidean distance and by 0.29% and 0.55%, respectively, compared to those of IMED. Compared to cosine dissimilarity and cosine dissimilarity on the IMED standardized transforms, AIMED-D outperforms these metrics by 0.95% and 0.25%, respectively, and AIMED-C outperforms them by 1.37% and 0.67%, respectively.

From Table 2, we can see that the proposed AIMEDs can reduce the number of support vectors substantially. AIMED-C uses only 53% and 65% support vectors in comparison with traditional Euclidean distance and IMED, respectively. We attribute this sharp decrease in the number of support vectors to the fact that AIMED reflects the data distribution more accurately than traditional Euclidean distance dose. Traditional Euclidean distance is sensitive to noise and small deformation. The images that are similar to one another may exhibit sparse distribution in a Euclidean space, most of which must be treated as support vectors to guarantee satisfactory classification accuracy. On the contrary, these images are gathered together in the AIMED space, and only representative images need to be treated as support vectors. Therefore, the number of support vectors can be greatly reduced. In addition, fewer support vectors can save hardware memory requirements and speed up the training and recognition processes. As listed in Table 2, the training time and the test time with AIMED-C are only 47% and 54% of those of the Euclidean distance method, respectively.

4.2. Face recognition

In this experiment, we used the UMIST face database [24], which is a multi-view database consisting of 575 gray-scale images of 20 subjects. Each of the subjects covers a wide range of poses from profile to frontal views as well as a variety of races, genders, and appearances. In our experiment, all the images were scaled to the size of 56 × 46 pixels. A total of 10 images per person were randomly chosen as the training data set, and the remaining 375 images were used to form the test set. PCA [29] based on different distance definitions was used as the dimension reduction method. The cosine dissimilarity was not considered since it cannot be embedded in the

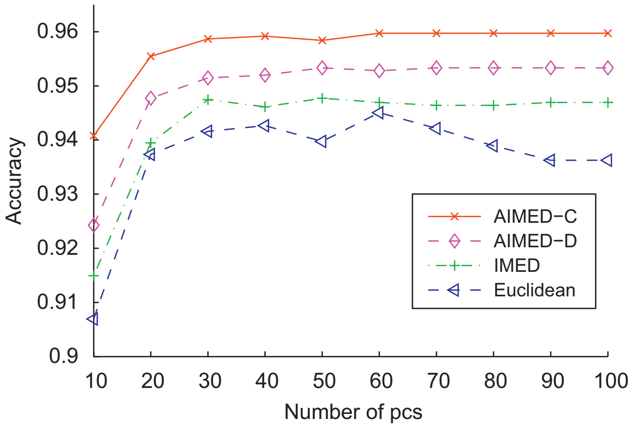

Fig. 5. Classification accuracies under different numbers of principal components in the UMIST database.

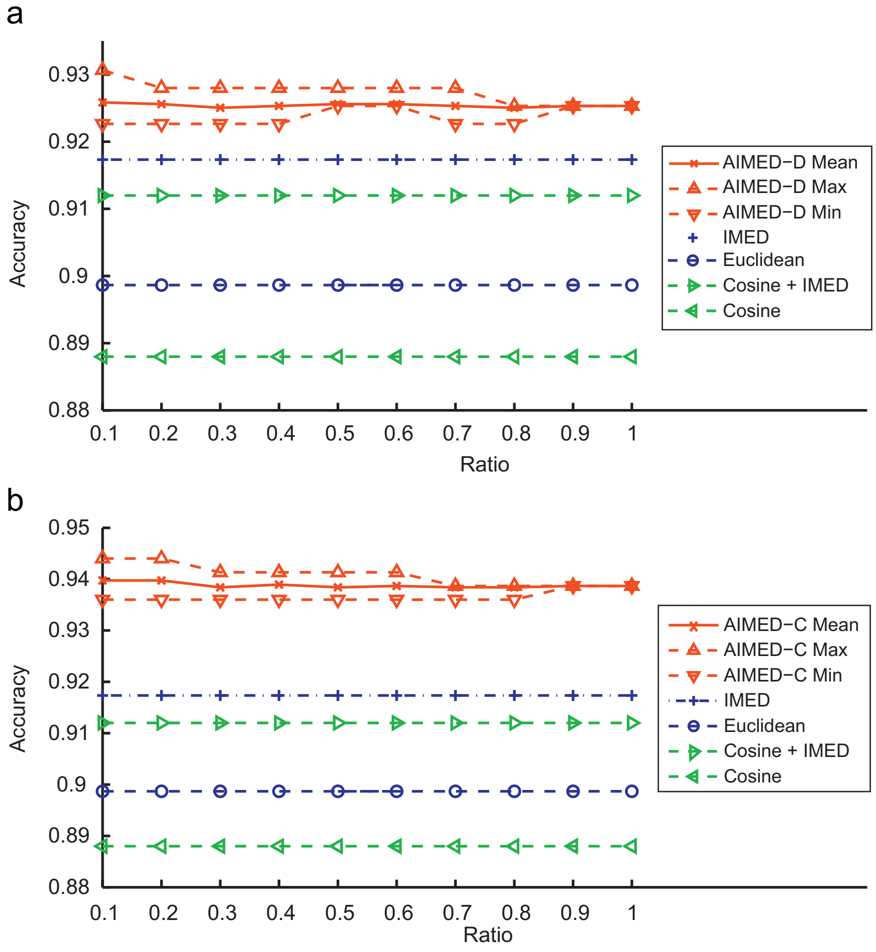

Fig. 6. Classification accuracies at different sampling ratios on the UMIST database: (a) AIMED-D and (b) AIMED-C.

PCA. After the dimension reduction, the nearest neighbor classifier was used to recognize faces. We repeated the experiments 10 times by randomly dividing the training and test data set, and the average classification accuracies under different numbers of principal components are depicted in Fig. 5.

From this figure, we can see that the two proposed AIMED methods always outperform both the traditional Euclidean distance and IMED under different numbers of principal components. The average accuracy improvement for AIMED-D with respect to Euclidean distance and IMED is, respectively, 1.28% and 0.66%, and for AIMED-C, these values are 2.04% and 1.42%, respectively. Even when the number of principal components is as few as 10, the classification accuracy of AIMED-C can still be as high as 94.08%, which is better than the highest classification accuracy of Euclidean distance. In addition, the standard derivations of classification accuracy for AIMED-C, AIMED-D, IMED, and Euclidean distance are 0.59%, 0.90%, 1.01%, and 1.09%, respectively. This indicates that AIMED-C has the most stable performance among the four distance metrics.

The matrices G_D and G_C were calculated on all the training data in the preceding experiments. We have also carried out experiments using parts of training data to estimate G_D and G_C. The nearest neighbor algorithm was used as the classifier to recognize faces in this experiment. We carried out experiments at the sample ratio from 10% to 100%. At each sample ratio, images were randomly chosen from the training data set 10 times, and G_D and G_C were calculated on the selected images for each time. The mean classification accuracies at different sample ratios are shown in Fig. 6. The best and the worst performances at different sample ratios are also shown in this figure. The results of using traditional Euclidean distance and IMED are also shown as the baseline. Since the nearest neighbor classifier was used in the experiments, we also show the results of cosine dissimilarity and cosine dissimilarity between the IMED standard transformed images for comparison. We can see that the proposed AIMEDs outperform all the other distances, even though only a small portion of the training data is used to estimate the matrices G_D and G_C.

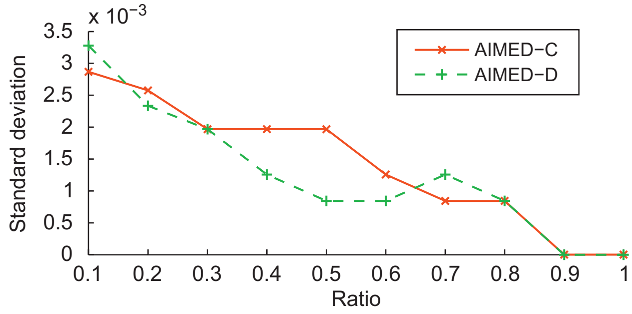

Fig. 7. Standard deviation of the accuracies at different sampling ratios on the UMIST database.

We also find that the performance of the AIMED is very robust. The best and the worst performances at each sample ratio are very close to each other. We calculated the standard deviation of the classification accuracies at each sampling ratio, which is shown in Fig. 7. We can see that the deviations are very small. The highest values of AIMED-D and AIMED-C are 0.0029 and 0.0033, respectively.

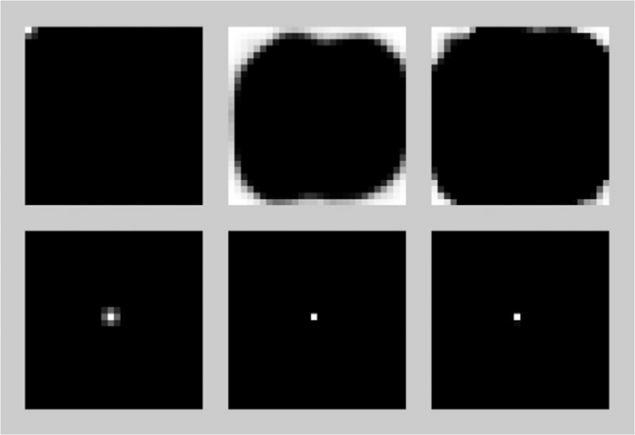

Fig. 8. Masks generated by different distance metrics at different locations. The first row: masks at the location of the upper left corner; the second row: masks at the location of the center. The values of $G^{1/2}$, $G_D^{1/2}$, and $G_C^{1/2}$ are shown from left to right in each row, respectively.

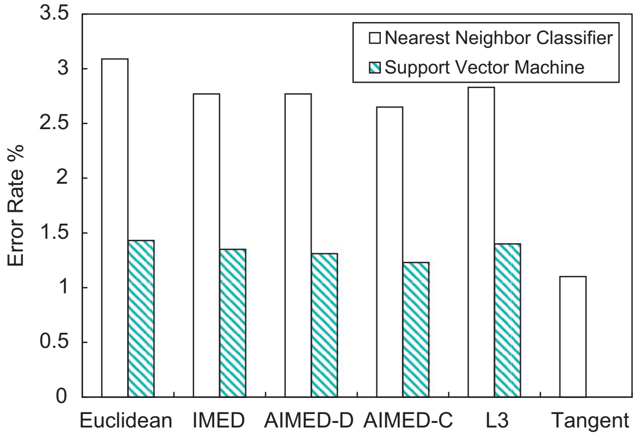

Fig. 9. Error rates with different image distance metrics on the MNIST database.

On the whole, the standard deviation decreases with the increasing of sample ratio.

From Figs. 4(a) and 6, we find that the cosine dissimilarity achieves comparable performance with Euclidean distance. We think the reason is that cosine dissimilarity calculates the angle between vectors just in the same space as Euclidean distance. While AIMED-C decorrelates the relationship between pixels, it computes the distance between images in a different space from that of Euclidean distance.

4.3. Handwritten digital recognition

To further evaluate the proposed distance metrics, the handwritten digital images from the MNIST [25] database were used in our experiments. This database contains a training set of 60,000 images, with a test set of 10,000 images. In this experiment, we used the nearest neighbor classifier and SVMs with radial basis functions to recognize digits with different distance metrics. The values of $G^{1/2}$, $G_D^{1/2}$, and $G_C^{1/2}$ at different locations are shown in Fig. 8. From this figure, we can see that $G_D^{1/2}$ and $G_C^{1/2}$ at the upper left corner reflect the background of digits. The large difference between pixels located on the background is apportioned to some smaller differences on the background by the mask, and the distance will be decreased. On the other hand, digital images of different numbers are highly distinct, therefore $G_D^{1/2}$ and $G_C^{1/2}$ at the center point almost have a single nonzero value.

The classification results with different image distance metrics on this database are shown in Fig. 9. From this figure, we can see that AIMED-D achieves comparable performance to IMED, and AIMED-C achieves better performance than IMED and Euclidean distance, both embedded in nearest neighbor classifier and in SVMs. There is almost no gray level deformation in this database since the gray level values of the images are almost '0' or '255'. Therefore, the effect of AIMED in handwritten images is less pronounced than that in face images.

The classification results of using the L3 distance and tangent distance [5] are also shown in Fig. 9 for comparison. The tangent distance was developed specifically for handwritten digit recognition, and has achieved the best performance among image similarity measures on this database [25]. From Fig. 9, we can see that the AIMED with nearest neighbor classifiers performs better than the L3 distance, but cannot perform as well as tangent distance. However, AIMED can be easily embedded in other pattern recognition algorithms, and benefit from the high classification accuracies of these algorithms. In this experiment, the performance of AIMED-C with SVMs is very close to that of tangent distance, which is the best performer of image similarity measures in this database.

5. Conclusions

With the characteristic of considering spatial relationship between pixels and the ability of being easy embedded in existing pattern recognition algorithms, IMED is a preferred distance measuring method for images. Based on IMED, we have proposed AIMED, which considers not only the spatial relationship between pixels, but also the gray level relationship between pixels. Our proposed AIMED makes the metric matrix adaptable to the content of the images and can reflect the shapes of the images. The experimental results on embedding AIMED into nearest neighbor classifiers, principal component analysis (PCA), and support vector machines (SVMs) demonstrate that the proposed AIMED achieves higher classification accuracy than both traditional Euclidean distance and the original IMED. Moreover, the proposed AIMED can gather similar images and reduce the number of support vectors when it is embedded in SVMs, and the performance of AIMED is more stable than that of the traditional Euclidean distance and IMED when it is embedded in PCA.

Acknowledgments

The authors would like to thank Zhi-Gang Fan, Bin Xia, Wen-Yun Yang, and Yang Yang for their help and discussions during the preparation of this paper, and the reviewers for their constructive comments and useful suggestions of the manuscript. This work was partially supported by the National Natural Science Foundation of China under the Grants NSFC 60375022, NSFC 60473040, and NSFC 60773090, the MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Shanghai Jiao Tong University, and the Okawa Foundation Research Grant.

References

[1] A. Tversky, I. Gati, Similarity, separability, and the triangle inequality, Psychol. Rev. 89 (2) (1982) 123–154.

[2] H.S. Seung, D.D. Lee, COGNITION: the manifold ways of perception, Science 290 (5500) (2000) 2268–2269.

[3] I.H. Witten, T.C. Bell, A. Moffat, Managing Gigabytes: Compressing and Indexing Documents and Images, Morgan Kaufmann, Los Altos, CA, 1999.

[4] C.C. Aggarwal, A. Hinneburg, D. Keim, On the Surprising Behavior of Distance Metrics in High Dimensional Space, Springer, Berlin, 2000.

[5] P. Simard, Y. LeCun, J.S. Denker, Efficient pattern recognition using a new transformation distance, Adv. Neural Inf. Process. Syst. (1992) 50–58.

[6] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using the Hausdorff distance, IEEE Trans. Pattern Anal. Mach. Intell. 15 (9) (1993) 850–863.

[7] E.P. Vivek, N. Sudha, Robust Hausdorff distance measure for face recognition, Pattern Recognition 40 (2) (2007) 431–442.

[8] C.H.T. Yang, S.H. Lai, L.W. Chang, Hybrid image matching combining Hausdorff distance with normalized gradient matching, Pattern Recognition 40 (4) (2007) 1173–1181.

[9] S. Santini, R. Jain, Similarity measures, IEEE Trans. Pattern Anal. Mach. Intell. 21 (9) (1999) 871–883.

[10] D. Jacobs, D. Weinshall, Y. Gdalyahu, Classification with nonmetric distances: image retrieval and classrepresentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (6) (2000) 583–600.

[11] X. Tan, S. Chen, J. Li, Z.H. Zhou, Learning non-metric partial similarity based on maximal margin criterion, in: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2006, pp. 168–175.

[12] J.B. Tenenbaum, V. Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.

[13] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.

[14] L. Wang, Y. Zhang, J. Feng, On the Euclidean distance of images, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1334–1339.

[15] J. Chen, R. Wang, S. Shan, X. Chen, W. Gao, Isomap based on the image Euclidean distance, in: Proceedings of the 18th International Conference on Pattern Recognition, vol. 2, 2006, pp. 1110–1113.

[16] K. Han, X.C. Zhu, Research on face recognition based on IMED and 2DPCA, J. Electron. 23 (5) (2006) 786–790.

[17] X. Mei, S.K. Zhou, H. Wu, Integrated detection, tracking and recognition for IR video-based vehicle classification, IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 5, 2006, pp. 745–748.

[18] T. Tangkuampien, D. Suter, 3D object pose inference via kernel principal component analysis with image Euclidian distance (IMED), in: Proceedings of the 2006 British Machine Vision Association, vol. 1, 2006, pp. 137–146.

[19] R. Wang, J. Chen, S. Shan, W. Gao, Enhancing training set for face detection, in: Proceedings of the 18th International Conference on Pattern Recognition, vol. 3, 2006, pp. 477–480.

[20] J. Li, B.L. Lu, A framework for multi-view gender classification, in: Proceedings of the 14th International Conference on Neural Information Processing, Lecture Notes in Computer Science, vol. 4984, Springer, Berlin, 2007, pp. 973–982.

[21] J.M. Lee, Riemannian Manifolds: An Introduction to Curvature, Springer, Berlin, 1997.

[22] J. Jost, Riemannian Geometry and Geometric Analysis, Springer, Berlin, 1995.

[23] W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang, D. Zhao, The CAS-PEAL large-scale Chinese face database and baseline evaluations, Technical report of JDL ⟨http://www.jdl.ac.cn/peal/pealtr.pdf⟩.

[24] D. Graham, N. Allinson, Characterizing virtual eigensignatures for general purpose face recognition, face recognition: from theory to applications, NATO ASI Series F, Computer and Systems Sciences 163 (1998) 446–456.

[25] Y. LeCun, The MNIST database of handwritten digits, Available online at: ⟨http://yann.lecun.com/exdb/mnist/index.html⟩.

[26] B. Moghaddam, M.H. Yang, Learning gender with support faces, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 707–711.

[27] H.C. Lian, B.L. Lu, Multi-view gender classification using multi-resolution local binary patterns and support vector machines, Int. J. Neural Syst. 17 (6) (2007) 479–487.

[28] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines ⟨http://www.csie.ntu.edu.tw/ cjlin/libsvm⟩.

[29] M. Turk, A. Pentland, Face recognition using eigenfaces, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586–591.

About the Author—JING LI received her B.S. degree in 1997 from Northeast University, China. She is currently a Ph.D. student at the Department of Computer Science and Engineering, Shanghai Jiao Tong University (SJTU), China. Her research interests include pattern recognition, computer vision, and incremental learning.

About the Author—BAO-LIANG LU is a professor of Computer Science and Engineering at Shanghai Jiao Tong University (SJTU). He received his B.S. degree in instrument and control engineering from Qingdao University of Science and Technology, China, in 1982, the M.S. degree in computer science and engineering from Northwestern Polytechnical University, China, in 1989, and the Dr. Eng. degree in electrical engineering from Kyoto University, Japan, in 1994. From 1982 to 1986, he was with the Qingdao University of Science and Technology. From April 1994 to March 1999, he was a Frontier Researcher at the Bio-Mimetic Control Research Center, the Institute of Physical and Chemical Research (RIKEN), Japan. From April 1999 to August 2002, he was a Research Scientist at the RIKEN Brain Science Institute. Since August 2002, he has been a full Professor at the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. His research interests include brain-like computing, neural network, machine learning, pattern recognition, brain–computer interface, computational linguistics, and computational biology and bioinformatics. He is a senior member of the IEEE.