Manifold-Manifold Distance with Application to Face Recognition based on Image Set

Ruiping Wang[1, 2, 3], Shiguang Shan[1, 2], Xilin Chen[1, 2], Wen Gao[4, 2]

[1]Key Lab of Intelligent Information Processing, Chinese Academy of Sciences (CAS), Beijing, China
[2]Digital Media Research Center, Institute of Computing Technology, CAS, Beijing, 100190, China
[3]Graduate School of the Chinese Academy of Sciences, Beijing, 100039, China
[4]School of EE&CS, Peking University, Beijing, 100871, China
{rpwang, sgshan, xlchen, wgao}@jdl.ac.cn

Abstract

In this paper, we address the problem of classifying image sets, each of which contains images belonging to the same class but covering large variations in, for instance, viewpoint and illumination. We innovatively formulate the problem as the computation of Manifold-Manifold Distance (MMD), i.e., calculating the distance between nonlinear manifolds each representing one image set. To compute MMD, we also propose a novel manifold learning approach, which expresses a manifold by a collection of local linear models, each depicted by a subspace. MMD is then converted to integrating the distances between pair of subspaces respectively from one of the involved manifolds.

The proposed MMD method is evaluated on the task of Face Recognition based on Image Set (FRIS). In FRIS, each known subject is enrolled with a set of facial images and modeled as a gallery manifold, while a testing subject is modeled as a probe manifold, which is then matched against all the gallery manifolds by MMD. Identification is achieved by seeking the minimum MMD. Experimental results on two public face databases, Honda/UCSD and CMU MoBo, demonstrate that the proposed MMD method outperforms the competing methods.

1. Introduction

In traditional visual recognition task, objects of interest are trained and recognized from only a few samples. However, with the increase of available video cameras and large capacity storage media, many new applications are emerging in which the image quantity of each object of interest for both training and testing can be very large. For example, as shown in Fig.1, nowadays, in many face recognition applications, a great number of images for each known subject have been able to be collected from video sequences or photo album, and recognition can also be conducted with a set of probe images rather than single probe image. In other words, recognition can be formulated as matching a probe image set against all the gallery image sets each representing one subject. We call this category of visual tasks as Object Recognition based on Image Set (ORIS) problem.

Over the past decade, ORIS problem has attracted increasing interest in computer vision community. However, it is worth pointing out that video-based object recognition is only a special case of ORIS. In ORIS, the images in the gallery or probe sets are collected not necessarily from consecutive video sequences but possibly from unordered photo album. Nevertheless, the images in the sets are generally of large amount and cover a variety of variations in the object's appearance, due to the camera pose changes, non-rigid deformations or different lighting conditions. Therefore, we can assume that the images in each image set distribute on a nonlinear manifold. Thus, this kind of ORIS problems can be converted to the problem of matching different manifolds, which is the basic idea of this paper.

In this paper, we formulate the ORIS problem as the computation of Manifold to Manifold Distance (MMD), i.e., calculating the distance between the gallery manifold learned from the training gallery image set and the probe manifold learned from the probe image set. We also propose a novel manifold learning approach, which expresses a manifold by a collection of local linear models, each depicted by a subspace. The MMD is then converted to integrating the distances between pair of subspaces respectively from one of the involved manifolds. In short, the main contributions of the paper lie in:

1) We formally and explicitly formulate the ORIS problem as the computation of what we name manifold to manifold distance (MMD).
2) After an overview on the distance measures on points, subspaces and manifolds, we propose a formal definition of MMD.
3) To facilitate the computation of MMD, a novel manifold learning method is proposed, which represents a nonlinear manifold as a collection of linear subspaces. Thus, MMD is converted to the integration of distances between pair of subspaces.
4) We define a more reasonable subspace distance, which measures not only the dissimilarity between the data variation modes of two subspaces but also the dissimilarity of the data itself.
5) The proposed MMD method is applied to Face Recognition based on Image Set (FRIS) problem, and impressive results are achieved.

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

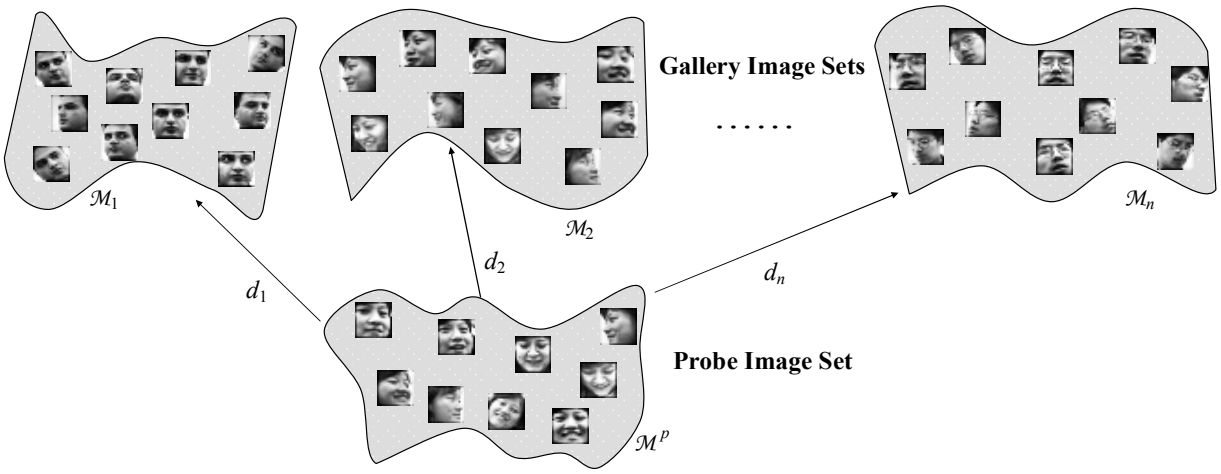


Figure 1: Face Recognition based on Image Set (FRIS). In FRIS, each subject is enrolled with a gallery image set, and the unknown subject is also represented by a probe image set. Note that the example facial images are from Honda/UCSD database [12].

2. Related work

Recently, there is growing interest in designing novel methods for multiple shots based object recognition [10], [12], [13], [15], [20], [24], [28], [29], [30]. Many of them are video-based. However, since this paper mainly focuses on problems in which the samples in the image sets are unordered, those previous works using video dynamics [12], [14], [15], [20], [31] are less mentioned in this section.

Broadly speaking, relevant approaches to ORIS problem mainly fall into two categories [10]: parametric model based and nonparametric sample based. Representative parametric methods include [1], [19]. They tend to represent each image set by a parametric distribution function and then measure the similarity between two distributions in terms of the Kullback-Leibler Divergence (KLD). The main drawbacks of parametric methods are that they need to solve the difficult parameter estimation problem and may easily fail when the training and novel test data sets have weak statistical relationships. Due to these limitations, several non-parametric methods are proposed based on matching pair-wise samples in the image sets. The basic premise behind is that the similarity between two sets can be reflected by the similarity of their corresponding modes (NN samples). For example, Hadid et al.[7] recently propose to use representative samples, called "exemplar", for image-based matching. By representing the image set as a nonlinear manifold, they extract such exemplars from the manifold using the Locally Linear Embedding (LLE) [18] and *k*-means clustering method.

More recently, by representing the image set as a linear subspace spanned by the images, the ORIS problem is converted to measuring the similarity or distance between subspaces. Representative methods include [10], [17], [27], [28], following the early works in [3], [8]. Basically, all these methods exploit the principal angles as their building blocks. The angles between two subspaces, which mainly reflect the common modes of variation of the two subspaces, are used as a similarity measure. As a method for comparing sets, the main advantages of principal angles include accuracy, efficiency, and robustness.

In spirit, the proposed MMD bears some resemblance to [4], [7], [12], [24] in that they also represent the image set by nonlinear manifold. It also shares the common idea of exploiting principal angles as distance measure with [9], [10], [27], [28]. However, our method has significant differences from these methods. In the next sections, we will formulate the problem more exactly and then present details of different components of our method.

3. Problem formulation

For visual object recognition, patterns can be represented in three possible levels: *point* (i.e., individual sample), *subspace* (i.e., linear model spanned by some samples), and *manifold* (i.e., nonlinear low-dimensional embedding learned from a large number of samples). We believe, in some sense, the core of pattern classification is the distance computation among these representations. The distance over points and subspaces has been well investigated in the literature. However, to our knowledge, very few studies have been done on the distance measure for manifolds.

In this section, after an overview of existing distance measures on points and subspaces, we give a primary formulation of manifold to manifold distance (MMD).

As illustrated in Fig. 2, the distance over points and subspaces include: point to point distance (PPD), point to subspace distance (PSD), and subspace to subspace distance (SSD). Note that hereinafter we always denote points by x_i, y_i, subspaces by S_i, and manifolds by $\mathcal{M}_i$.

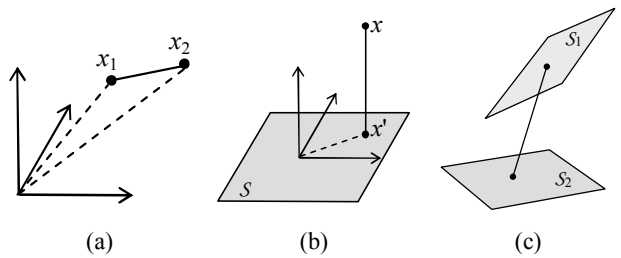


Figure 2: Three types of distances defined over points and subspaces. (a) Point to point distance (PPD). (b) Point to subspace distance (PSD). (c) Subspace to subspace distance (SSD).

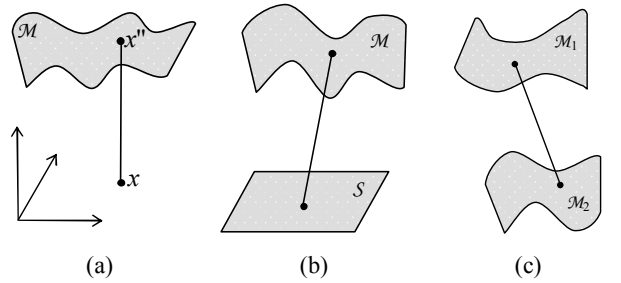


Figure 3: Three types of distances defined over manifolds. (a) Point to manifold distance (PMD). (b) Subspace to manifold distance (SMD). (c) Manifold to manifold distance (MMD).

Point to point distance (PPD): denote by $d(x_1, x_2)$ the distance from point x_1 to x_2. The most commonly used PPD is the traditional Euclidean distance as follows:

$$d(x_1, x_2) = \|x_1 - x_2\|. \quad (1)$$

Point to subspace distance (PSD): denote by $d(x, S)$ the distance from point x to subspace S. It is generally defined as the so-called L2-Hausdorff distance:

$$d(x, S) = \min_{y \in S} \|x - y\| = \|x - x'\|. \quad (2)$$

In fact, x' is the projection of x in the subspace S, also the nearest point to x in S. Thus, the PSD is actually the PPD from x to its projection x' in S. It is also known as "distance-from-feature-space" (DFFS) in [16], [22].

Subspace to subspace distance (SSD): denote by $d(S_1, S_2)$ the distance between two subspaces S_1 and S_2. To our knowledge, there is not a unified definition yet to measure the SSD. Perhaps, the concept of principal angles is the most commonly exploited one due to its favorable performance [10], [17], [27], [28]. Recently, another SSD is proposed in [26], which can be regarded as utilizing the sum of DFFS between the bases of two subspaces.

Manifold learning has become a very active research area in computer vision, pattern recognition and machine learning. Typically, the underlying structure of high-dimensional observation samples whose variations are controlled by only a few factors can be modeled by a low-dimensional manifold [18], [21]. While many methods have been developed to compute the manifold embedding, to our knowledge, the topic of a general distance measure over nonlinear manifolds has not been given proper attention in the literature. In the following part of this section, we make a primary attempt to define distances over points, subspaces, and manifold. There are also three categories of distances: point to manifold distance (PMD), subspace to manifold distance (SMD), and manifold to manifold distance (MMD), which are also shown in Fig. 3.

Our main motivation comes from the fact that local linearity property holds everywhere on a global nonlinear manifold, and thus manifold can be modeled by a collection of local linear subspaces [18]. Therefore, the distances associated with manifold can be converted to those defined on subspaces. In general, the manifold to manifold distance can be viewed as extending subspace distance to account for more general and complex data variations. Formally, we denote the component subspaces of a manifold $\mathcal{M}$ by C_i, and express a manifold as:

$$\mathcal{M} = \{C_1, C_2, \ldots, C_m\} \quad (3)$$

where m is the number of local linear subspaces.

Point to manifold distance (PMD): denote by $d(x, \mathcal{M})$ the distance from point x to manifold $\mathcal{M}$. Similar to "point to subspace" distance, one can define this distance by finding the closest point to x in $\mathcal{M}$ as follows:

$$d(x, \mathcal{M}) = \min_{C_i \in \mathcal{M}} d(x, C_i) = \min_{C_i \in \mathcal{M}} \min_{y \in C_i} \|x - y\| = \|x - x''\|. \quad (4)$$

In analogy to x' in the PSD, here we call x'' the projection of x in the manifold $\mathcal{M}$.

Subspace to manifold distance (SMD): denote by $d(S, \mathcal{M})$ the distance from subspace S to manifold $\mathcal{M}$. It can be defined by seeking the closest subspace to S in manifold $\mathcal{M}$:

$$d(S, \mathcal{M}) = \min_{C_i \in \mathcal{M}} d(S, C_i). \quad (5)$$

It comes that, SMD is reduced to SSD. In the next section, a novel and more reasonable SSD function will be described.

Manifold to manifold distance (MMD): denote by $d(\mathcal{M}_1, \mathcal{M}_2)$ the distance between manifold $\mathcal{M}_1$ and manifold $\mathcal{M}_2$. Before giving its function definition, let us recall the FRIS problem shown in Fig. 1. Typically, when the gallery and probe image sets belonging to the same subject contain images taken from different views but with a certain overlap in views, global data characteristics of the two sets might be very different. So, to match the two sets as the same class, the most effective solution is to find the common views and measure the similarity of those parts of data. Therefore, we define MMD by the closest subspace pair from the two manifolds as follows:

$$d(\mathcal{M}_1,\mathcal{M}_2)=\min_{C_i\in\mathcal{M}_1} d(C_i,\mathcal{M}_2)=\min_{C_i\in\mathcal{M}_1}\min_{C_j'\in\mathcal{M}_2} d(C_i,C_j')\,. \qquad (6)$$

Clearly, the similarity between two manifolds is computed as the similarity between their best suited local models.

It is worth noting that, to compute MMD, the proposed measure in Eq. (6) is just one of the possibilities. While many others may be explored, however, for the ORIS task, Eq. (6) is believed to be one of the most appropriate.

4. Computing Manifold to Manifold Distance

As described above, we express the nonlinear manifold as a collection of local linear models, and then integrate the subspace distances to yield the final MMD. First, we give a brief introduction of principal angles, which serves as an important ingredient of our local model similarity measure. Then, we propose our local model constructing method. By defining a more reasonable local model similarity, we finally derive the MMD.

4.1. Principal angles

Principal angles $0\le\theta_1\le\theta_2\le\ldots\le\theta_r\le(\pi/2)$ between two linear subspaces S_1 and S_2 are uniquely defined as:

$$\begin{aligned}&\cos(\theta_k)=\max_{\boldsymbol{u}_k\in S_1}\max_{\boldsymbol{v}_k\in S_2}\boldsymbol{u}_k^T\boldsymbol{v}_k\\ &s.t.\quad \boldsymbol{u}_k^T\boldsymbol{u}_k=\boldsymbol{v}_k^T\boldsymbol{v}_k=1,\ \boldsymbol{u}_k^T\boldsymbol{u}_i=\boldsymbol{v}_k^T\boldsymbol{v}_i=0,\\ &\qquad i=1,2,\ldots,k-1\end{aligned} \qquad (7)$$

where $r=\min(\dim(S_1),\dim(S_2))$ [8]. $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$ are called the k-th pair of canonical vectors. The cosines of the principal angles are called canonical correlations.

A numerically stable algorithm to compute the principal angles was proposed in [3] based on Singular Value Decomposition (SVD). Let $\mathbf{P}_1\in\mathbb{R}^{D\times d_1}$ and $\mathbf{P}_2\in\mathbb{R}^{D\times d_2}$ respectively denote the orthonormal basis of two subspaces S_1 and S_2, d_1 and d_2 are the subspace dimension. The SVD of $\mathbf{P}_1^T\mathbf{P}_2$ is as follows:

$$\mathbf{P}_1^T\mathbf{P}_2=\mathbf{Q}_{12}\mathbf{\Lambda}\mathbf{Q}_{21}^T \quad s.t. \quad \mathbf{\Lambda}=diag(\sigma_1,\ldots,\sigma_r) \qquad (8)$$

where $\mathbf{Q}_{12}$ and $\mathbf{Q}_{21}$ are orthogonal matrices. The singular values $\sigma_1,\ldots,\sigma_r$ are the cosines of the principal angles, i.e. canonical correlations:

$$\cos(\theta_k)=\sigma_k,\quad k=1,2,\ldots,r\,. \qquad (9)$$

The associated canonical vectors are $\mathbf{U}=\mathbf{P}_1\mathbf{Q}_{12}=[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{d_1}]$, $\mathbf{V}=\mathbf{P}_2\mathbf{Q}_{21}=[\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{d_2}]$. If the maximum principal angle is small, the subspaces are close to each other. Intuitively, the first pair of canonical vectors corresponds to the most similar modes of variation of two linear subspaces; every next pair to the most similar modes orthogonal to all previous ones.

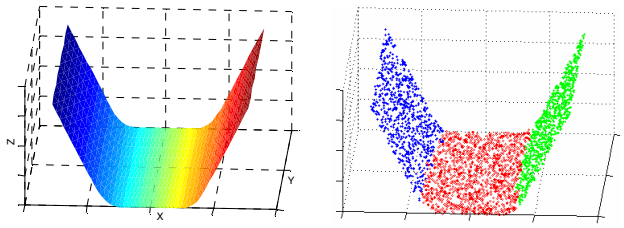


Figure 4: *Left*: 3D "U-like shape" manifold, which is patch-wise linear with three planes smoothly connected with each other. *Right*: Three MLPs (encoded with distinct *colors*) constructed by our method. Each MLP is then modeled by a linear subspace.

4.2. Local linear model construction

To construct local linear models from nonlinear manifold, previous work has presented several approaches [4], [7], [9], [12]. However, they typically use iterative-based clustering methods such as k-means to assign samples into different clusters. This procedure has two main limitations: first, the number of target clusters need to be specified a priori manually; second, the linearity property of the extracted local models is not guaranteed explicitly. To overcome these problems, we develop an efficient one-shot algorithm for adaptively constructing local linear models. Compared with previous methods, our algorithm can effectively guarantee the linear property of the local models and meanwhile gains more computational efficiency.

We first introduce a reasonable and compact definition of local linear patch on the manifold, called Maximal Linear Patch (MLP). Inspired by geometric intuition, MLP is defined to span a maximal linear subspace and its linear perturbation is naturally reflected by the deviation between the Euclidean distances and geodesic distances [21] in the patch, which is more tractable according to exactly designed constraints. See Fig. 4 for a conceptual illustration. By this new concept of local linear patch, the basic idea of our one-shot clustering method is that, each new MLP is stemmed from a seed point gradually until the linearity constraint is broken. This procedure can effectively guarantee the local linear property and adaptively control the number of local models.

Formally, a data set $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_N\}$ is given, where $\boldsymbol{x}_i\in\mathbb{R}^D$ is a D-dimensional column vector, and N is the sample number. The samples in the set are assumed to come from a low-dimensional manifold $\mathcal{M}$. We aim to partition the data set $\boldsymbol{X}$ to a collection of disjoint MLPs, i.e., local models C_i. That is,

$$\begin{aligned}&\boldsymbol{X}=\bigcup_{i=1}^{m}C_i\,,\\ &C_i\cap C_j=\varnothing\ \ (i\ne j,\ i,j=1,2,...,m),\\ &C_i=\{\boldsymbol{x}_1^{(i)},\boldsymbol{x}_2^{(i)},...,\boldsymbol{x}_{Ni}^{(i)}\}\ \ (\sum_{i=1}^{m}Ni=N),\end{aligned} \qquad (10)$$

where m is the number of patches and Ni is the number of points in patch C_i.

To perform the partitioning, both pair-wise Euclidean distance matrix $\boldsymbol{D}_E$ and geodesic distance matrix $\boldsymbol{D}_G$ [21], are obtained firstly. Then, the m MLPs are extracted using our one-shot method *Algorithm 1* as follows. Note that, the threshold parameter θ in Eq.(11) reflects the degree of linear perturbation of MLP. Easy to know, a larger θ implies fewer local models (thus more efficient) but larger linearity deviation, and vice versa. That is, θ controls the tradeoff between efficiency and accuracy.

Algorithm 1: Local model construction

1 Initialize that $i=1$, $C_i=\varnothing$, $\boldsymbol{X}_T=\varnothing$, $\boldsymbol{X}_R=\boldsymbol{X}$;

2 While ($\boldsymbol{X}_R \neq \varnothing$)

2.1 Randomly select a seed point from $\boldsymbol{X}_R$ as $\boldsymbol{x}_1^{(i)}$, update $C_i=\{\boldsymbol{x}_1^{(i)}\}$, $\boldsymbol{X}_R=\boldsymbol{X}_R-\{\boldsymbol{x}_1^{(i)}\}$;

2.2 For ($\forall\, \boldsymbol{x}_m^{(i)} \in C_i$)

Identify each of its *k*-NNs $\boldsymbol{x}_c$ as *candidate*. If $\boldsymbol{x}_c$ satisfies simultaneously $\boldsymbol{x}_c \in \boldsymbol{X}_R$ and

$$\boldsymbol{D}_G(\boldsymbol{x}_c,\boldsymbol{x}_n^{(i)}) / \boldsymbol{D}_E(\boldsymbol{x}_c,\boldsymbol{x}_n^{(i)}) < \theta \quad (\text{for } \forall\, \boldsymbol{x}_n^{(i)} \in C_i) \tag{11}$$

then update $C_i = C_i \cup \{\boldsymbol{x}_c\}$, $\boldsymbol{X}_R = \boldsymbol{X}_R - \{\boldsymbol{x}_c\}$;

2.3 Step 2.2 is stopped when there is no candidate point can be added into C_i.

2.4 $\boldsymbol{X}_T = \bigcup_{j=1}^{i} C_j$, $\boldsymbol{X}_R = \boldsymbol{X} - \boldsymbol{X}_T$; $i \leftarrow i+1$, $C_i=\varnothing$;

Discussion: On the linear criterion of MLP, we exploit the ratio between two distances of each data pair as Eq.(11). Alternative strategies such as the difference between two distances may also be used. We believe that these strategies are in some sense equivalent. Another feature of our algorithm is the sequential and one-shot manner to construct MLPs. Since it is non-iterative, high efficiency can be guaranteed. One problem is that, the sequential manner may benefit the patches earlier computed, and those computed later might have smaller size. Fortunately, this problem can be handled by using a simple post-processing strategy based on PCA, which combines small patches into larger ones that can better cover the intrinsic linear subspace.

To represent the local models, i.e. MLPs, we employ PCA subspace for its simplicity and efficiency. For each local model C_i, we denote its sample mean by $\boldsymbol{e}_i$ and the corresponding principal component matrix by $\mathbf{P}_i \in \mathbb{R}^{D\times d_i}$, which forms a set of orthonormal basis of the subspace. Here, the PCA dimension d_i is chosen to preserve 95% variances. Some local models constructed for two of the individuals in Fig.1 are shown in Fig. 5. It can be seen that, samples in a single local model exhibit slight appearance variations and they all look similar to the sample mean. Hereinafter, we call the sample centers "exemplars", which can represent the samples in the local models to some extent and serve as another important ingredient in our local model similarity measure. This will be clear in the next section.

Figure 5: Some local models. Each row shows a local model with the sample mean (first column), i.e., exemplar, and 6 representative samples. The 1st and 2nd rows belong to one individual, and the 3rd and 4th rows belong to another individual.

4.3. Manifold to manifold distance

By representing the local models by linear subspaces, we only need to define the "subspace to subspace" distance for the final MMD in Eq.(6). As discussed in Sec.3, some previous methods have utilized principal angles, which mainly reflect the common modes of variations between two subspaces while ignoring the data itself. We call their distance measure as "variation based measure". In another aspect, several methods [4], [7] have used the sample means in the local models to measure the local model similarity. We call their distance as "exemplar based measure". Since the subspace (or local model) is spanned by a set of samples, the two types of distance measures emphasize only either the similarity of data variation modes or the similarity of data samples itself. However, easy to understand, it is better to fuse both measures to give a more complete distance measure, which is just what we do in this paper.

For two subspaces C_i and C_j constructed above, we denote their corresponding exemplars and orthonormal bases by $\boldsymbol{e}_i$, $\boldsymbol{e}_j$ and $\mathbf{P}_i \in \mathbb{R}^{D\times d_i}$ $\mathbf{P}_j \in \mathbb{R}^{D\times d_j}$. We define the *variation distance measure* of the two subspaces by the average of canonical correlations as follows:

$$d_V(C_i,C_j) = r \cdot 1/trace(\boldsymbol{\Lambda}) = r/\sum_{k=1}^{r}\sigma_k \,. \tag{12}$$

where $\sigma_1,\ldots,\sigma_r$ are canonical correlations in Sec.4.1 and $r=\min(d_i,d_j)$. We then define the *exemplar distance measure* of the two subspaces by the correlation of the two exemplar samples:

$$d_E(C_i, C_j) = \| \boldsymbol{e}_i \| \cdot \| \boldsymbol{e}_j \| / \boldsymbol{e}_i^T \boldsymbol{e}_j . \quad (13)$$

Finally, our subspace distance measure comes in the form of a weighted average of the *variation distance measure* and *exemplar distance measure* as:

$$d(C_i, C_j) = (1-\alpha) \cdot d_E(C_i, C_j) + \alpha \cdot d_V(C_i, C_j) . \quad (14)$$

When applying to comparing two image sets, the two measures complement each other: the former describes how similar the appearance of images in the two sets, whereas the latter reflects how close the common variation modes of images in the two sets.

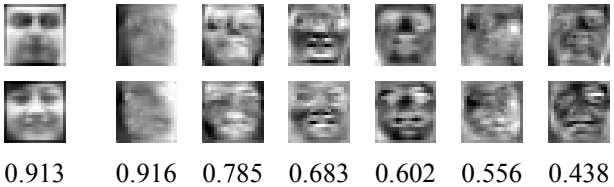
Now it is easy to compute the MMD we formulated in Eq. (6). Take the three manifolds $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}^p$ in Fig. 1 for example. By constructing local models for each manifold, we compute the distances over these manifolds. In Fig. 6(a) and (b), we respectively show the closest local model pair from $\mathcal{M}_1$ and $\mathcal{M}^p$ (the different individuals), and that from $\mathcal{M}_2$ and $\mathcal{M}^p$ (the same individual).

Discussion: For measuring invariant image similarity, two distances called Joint Manifold Distance (JMD) [5] and Multiresolution Manifold Distance (MRMD) [23] were proposed recently. While the titles seem similar to our MMD, the intrinsic properties of these two methods are very different from ours mainly in the following aspects:

1) Essentially, JMD and MRMD serve as distance measures between images to achieve invariance to parameterized image transformations; while MMD aims to measure the similarity between two sets of images from the nonparametric viewpoint.
2) Both JMD and MRMD are actually defined over points in linear subspaces; while MMD accounts for the distance of data variations on general manifolds.
3) Moreover, both JMD and MRMD involve iterative optimizations in a large parameter space; while our MMD is computed in a closed-form, which facilitates efficient online set matching applications.

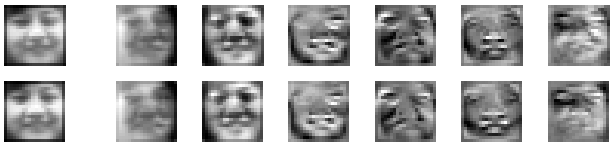
5. Experimental results

The application of the above principles to a hard real world problem offers some useful insights into our proposed framework. The application we consider is the task of Face Recognition based on Image Set (FRIS). In FRIS, each known subject is enrolled with a set of facial images and modeled as a gallery manifold, while a testing subject is modeled as a probe manifold, which is then matched against all the gallery manifolds by MMD. Identification is achieved by seeking the minimum MMD. Note that MMD is a general distance measure for set comparison and its computation does not incorporate any discriminative information. The following sections discuss the experimental results in detail.



0.913 0.916 0.785 0.683 0.602 0.556 0.438

(a)



0.998 0.987 0.983 0.966 0.951 0.891 0.849

(b)

Figure 6: Local model similarity. (a) Local model pair belongs to different individuals. (b) Local model pair belongs to the same individual. In both (a) and (b), the first column shows the exemplars of the two local models, and the rest columns show the first 6 canonical vectors in turn. From their correlation shown below the images, we can see that every pair of canonical vectors well captures similar variation modes.

5.1. Experimental setting

Datasets: We consider two different public datasets: Honda/UCSD [12] and CMU MoBo [6] in order to ensure an extensive evaluation of different methods against dataset changes including image resolution, facial expressions, illumination variations and the size of the dataset.

The first dataset, Honda/UCSD, is collected by Lee et al. for video-based face recognition. We consider their first subset, which has 59 video sequences of 20 different people (each person has at least 2 videos). Each video consists of about 300-500 frames. During the data collection, the individuals were asked to move in different combinations of rotation, expression, and speed. We use a cascaded face detector [25] to detect faces in each video sequence. Each faces are then resized to gray images of size 20×20, followed by histogram equalization to eliminate the lighting effects. Fig. 1 has shown some examples.

The second dataset, MoBo (Motion of Body), is the most commonly used in video-based face recognition research [7], [11], [15], [31]. It was originally collected for the purpose of human identification from distance. The considered subset contains 96 sequences of 24 different subjects walking on a treadmill (each person has 4 videos). Each sequence has 300 frames. From each video sequence, facial images are obtained in the same way as we did for the Honda dataset. The size of the resulted facial images is 30×30 pixels. Some example images are shown in Fig. 7.

For each individual of both datasets, one video sequence is used for training and the rest sequences are used for testing. On each dataset, we perform experiments for 5 randomly selected training/test combinations for reporting identification rates.

Figure 7: Examples of facial images from the MoBo database.

Comparative methods: We perform experiments on the following methods:

1. Nearest neighbor (NN) matching in (i) Eigenface, and (ii) Fisherface [2],
2. NN matching in LLE + *k*-means clustering [7], which is a typical exemplar-based method,
3. Mutual Subspace Method (MSM) [28], which is a typical variation-based method,
4. Our Manifold to Manifold Distance (MMD) method.

Parameter setting: In NN-Eigenface, the dimension is set to preserve 95% of data energy. In NN-Fisherface, PCA was applied first to avoid singularity problems and the dimension of LDA is set to the number of classes minus 1. In LLE + *k*-means, we use the same parameters setting as [7]. For each training video sequence, k=5 exemplars are extracted. These three methods all determine the identity of the probe sequence using majority voting scheme. In MSM, we first apply PCA to each image set to get its subspace basis. The PCA also preserves about 95% of data energy, and then the first 10 canonical correlations are exploited.

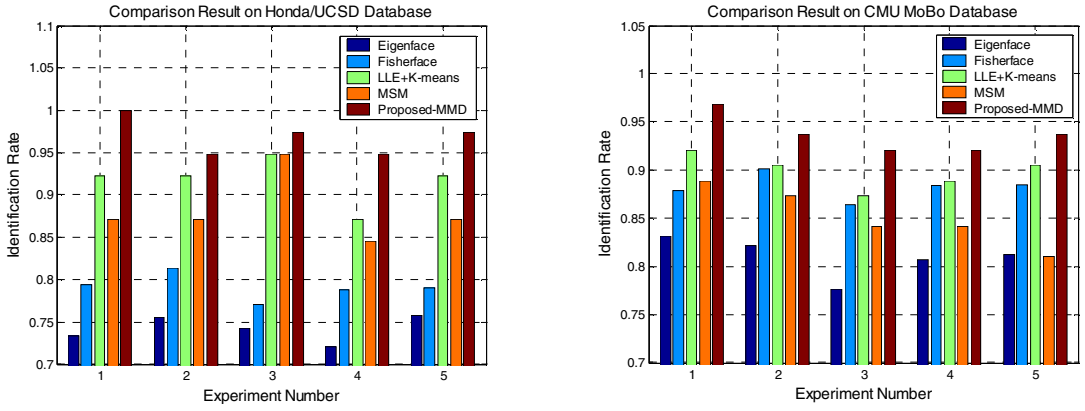
Our implementation: The important parameters in the proposed MMD include: (i) the threshold θ in Eq.(11). Under an empirical setting $\theta = 1.1$, one set with 300-400 images, which is typically the case in both datasets above, can obtain about 6-10 local models. (ii) The PCA dimension d_i for representing the local model C_i. By preserving 95% data variances in our work, the value of d_i varies from 5 to 10 for different local models. (iii) The weighting parameter α in Eq. (14). It is set to 0.5 for equal weights of the variation distance measure and exemplar distance measure.

5.2. Identification results and analysis

The identification performance for each 5 experiments on both databases is demonstrated in Fig. 8. The recognition rates shown in Table 1 are the average results over all 5 random trials of the evaluated algorithms.

First, it can be seen the two frame based methods – Eigenface and Fisherface yield relatively poor performance, especially in the Honda/UCSD database, which contains larger pose variations. Though it may seem not fair to compare them with the set based methods, the experiments suggest that frame based methods are more sensitive to large pose changes and may not work well in a real world unconstrained environment.

Second, the other three methods LLE + *k*-means, MSM and our MMD, though all are set based, yield different results due to respective properties. Among them, MSM exhibits the lowest recognition rates. This is not unexpected since MSM simply represents the complex image set as a linear subspace, and as a variation-based method, it only considers the similarity of variation modes of two subspaces. In contrast, both LLE + *k*-means and MMD model the facial image sets by nonlinear manifolds. However, their differences are also obvious. The LLE + *k*-means method mainly exploits the concept of manifold to extract representative samples, i.e., exemplars, from training sets. In the testing stage, images in each test set are classified separately, and the recognition problem is thus reduced to NN-type image matching similar to Eigenface. As noted in [4], this exemplar-based method is simple and may not fully characterize the variability of the image set. In contrast, the proposed MMD method treats both training and test sets as manifolds, and our local model distance function measures not only the similarity of data samples but also the similarity of data variation modes. By integrating the properties of MSM and LLE + *k*-means in a novel way, our MMD method hence yields significant performance benefits.



(a) (b)

Figure 8: Identification results of different algorithms on (a) Honda/UCSD database and (b) CMU MoBo database.

Table 1. Average recognition rates (%) of different methods on two public databases, Honda/UCSD and CMU MoBo.

	Honda/UCSD	CMU MoBo
Eigenface	74.2	81.0
Fisherface	79.2	88.3
LLE + *k*-means	91.8	89.8
MSM	88.2	85.1
Proposed MMD	**96.9**	**93.6**

6. Conclusion and future work

Comparing sets of images undergoing large variations is a challenging problem. By representing each image set by a nonlinear manifold, the problem is formulated as measuring the similarity between manifolds. To solve this problem, we propose a formal definition of Manifold to Manifold Distance (MMD), and present several technical contributions for its computation. The method is applied to face recognition based on image set. Extensive experimental results demonstrate that, even without any discriminative information, the proposed method still achieves better performance than the competing methods.

Currently the MMD is exploited mainly as a general set distance measure and we just propose one possible solution to its computation. In the future, we intend to incorporate discriminative information for supervised classification problems. Another interesting direction could be to learn the local model distance in more sophisticated manner.

Acknowledgements

This paper is partially supported by National Natural Science Foundation of China under contract No.60772071, No.60673091, and No.60728203; Hi-Tech Research and Development Program of China under contract No.2006AA01Z122 and No.2007AA01Z163; 100 Talents Program of CAS; and ISVISION Technology Co. Ltd.

References

[1] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face Recognition with Image Sets Using Manifold Density Divergence. *CVPR*, pp. 581–588, 2005.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *PAMI*, 19(7):711–720, 1997.

[3] Å. Björck and G. H. Golub. Numerical Methods for Computing Angles between Linear Subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

[4] W. Fan and D.-Y. Yeung. Locally Linear Models on Face Appearance Manifolds with Application to Dual-Subspace Based Classification. *CVPR*, pp. 1384–1390, 2006.

[5] A. W. Fitzgibbon and A. Zisserman. Joint Manifold Distance: A New Approach to Appearance based Clustering. *CVPR*, pp. 26–33, 2003.

[6] R. Gross and J. Shi. The CMU Motion of Body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June 2001.

[7] A. Hadid and M. Pietikäinen. From Still Image to Video-Based Face Recognition: An Experimental Analysis. *FG*, pp. 813–818, 2004.

[8] H. Hotelling. Relations between Two Sets of Variates. *Biometrika*, 28:321–372, 1936.

[9] T.K. Kim, O. Arandjelović, and R. Cipolla. Learning over Sets Using Boosted Manifold Principal Angles (BoMPA). *Proc. British Machine Vision Conf.*, pp. 779–788, 2005.

[10] T.K. Kim, J. Kittler, and R. Cipolla. Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations. *PAMI*, vol.29, no.6, pp.1005–1018, 2007.

[11] V. Krüeger and S. Zhou. Exemplar-based Face Recognition from Video. *ECCV*, volume 4, pp. 732–746, 2002.

[12] K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman. Video-Based Face Recognition Using Probabilistic Appearance Manifolds. *CVPR*, pp. 313–320, June 2003.

[13] Y. Li, S. Gong, and H. Liddell. Constructing Facial Identity Surface in A Nonlinear Discriminating Space. *CVPR*, volume 2, pp. 258–263, 2001.

[14] W. Liu, Z. Li, and X. Tang. Spatio-temporal Embedding for Statistical Face Recognition from Video. *ECCV*, 2006.

[15] X. Liu and T. Chen. Video-Based Face Recognition Using Adaptive Hidden Markov Models. *CVPR*, 2003.

[16] B. Moghaddam and A. Pentland. Probabilistic Visual Learning for Object Representation. *PAMI*, vol.19, 1997.

[17] M. Nishiyama, O. Yamaguchi, and K. Fukui. Face Recognition with the Multiple Constrained Mutual Subspace Method. *AVBPA*, pp. 71–80, 2005.

[18] S. Roweis and L. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, December 2000.

[19] G. Shakhnarovich, J. W. Fisher, and T. Darrell. Face Recognition from Long-term Observations. *ECCV*, 2002.

[20] J. Stallkamp, H.K. Ekenel, R. Stiefelhagen. Video-based Face Recognition on Real-World Data. *ICCV* 2007.

[21] J. Tenenbaum, V. Silva, and J. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323,December 2000.

[22] M. Turk and A. Pentland. Face Recognition Using Eigenfaces. *CVPR*, pp.586-591, 1991.

[23] N. Vasconcelos, and A. Lippman. A Multiresolution Manifold Distance for Invariant Image Similarity. *IEEE Trans. Multimedia*, vol. 7, no. 1, pp.127–142, 2005.

[24] V.Colin de Verdiére and J.L. Crowley. Visual Recognition Using Local Appearance. *ECCV*, vol. 1, pp. 640-654, 1998.

[25] P. Viola and M. Jones. Robust Real-Time Face Detection. *Int'l J. Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[26] L. Wang, X. Wang, J. Feng. Subspace Distance Analysis with Application to Adaptive Bayesian Algorithm for Face Recognition. *Pattern Recognition*,vol.39, pp.456–464,2006.

[27] L. Wolf and A. Shashua. Learning over Sets Using Kernel Principal Angles. *JMLR*, vol. 4, no. 10, pp. 913–931, 2003.

[28] O. Yamaguchi, K. Fukui, K. Maeda. Face Recognition Using Temporal Image Sequence. *FG*, pp. 318–323, 1998.

[29] J. Zhang, S.Z. Li, and J. Wang. Nearest Manifold Approach for Face Recognition. *FG,* pp. 223–228, 2004.

[30] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face Recognition:A Literature Survey. *ACM Comput. Surv*, 2003.

[31] S. Zhou and R. Chellappa. Probabilistic Human Recognition from Video. *ECCV*, volume 3, pages 681–697, 2002.