International Journal of Wavelets, Multiresolution and Information Processing

Face Recognition via Collaborative Representation based Multiple One-Dimensional Embedding

Yulong Wang, Yuan Yan Tang

Faculty of Science and Technology
University of Macau, Macau, 999078, China
wangyulong6251@gmail.com; yytang@umac.mo

Luoqing Li

Faculty of Mathematics and Statistics
Hubei University, Wuhan 430062, P. R. China
lilq@hubu.edu.cn

Jianzhong Wang

Department of Mathematics and Statistics
Sam Houston State University, Huntsville, TX 77341, USA
jzwang@shsu.edu

Received 19 March 2015
Revised 20 September 2015
Accepted 10 October 2015

This paper presents a novel classifier based on collaborative representation and multiple one-dimensional embedding with applications to face recognition. To use multiple 1-D embedding (1DME) framework in semi-supervised learning is first proposed by one of the authors, J. Wang, in 2014. The main idea of the multiple 1-D embedding is the following: Given a high-dimensional data set, we first map it onto several different 1-D sequences on the line while keeping the proximity of data points in the original ambient high-dimensional space. By this means, a classification problem on high dimension reduces to the one in a one-dimensional framework, which can be efficiently solved by any classical 1-D regularization method, for instance, an interpolation scheme. The dissimilarity metric plays an important role in learning a decent 1DME of the original data set. Our another contribution is to develop a collaborative representation based dissimilarity metric. Compared to the conventional Euclidean distance based metric, the proposed method can lead to better results. The experimental results on real-world databases verify the efficacy of the proposed method.

Keywords: Multiple 1-D embedding; smooth ordering; collaborative representation.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

Face recognition has received massive interest in pattern recognition for decades owing to its wide applications in the real world.[8, 11, 15, 28] It also has a rich literature.[4, 6, 9, 18, 26, 27] A number of face recognition algorithms have been developed to extract useful facial fea-

tures, including Fisherface,[2] Eigenface,[26] Independent Face,[14] and Laplacianface.[10] The existing literature[1,7] shows that facial images of a subject approximately lie in a low-dimensional subspace. To use this as a prior knowledge, Wright *et al.*[27] proposed a *sparse representation-based classifier* (SRC) by exploiting the discriminative property of sparse representation of the query image to perform classification. After the work of Wright *et al.*,[27] various representation-based classifiers are proposed for face recognition. For example, Zhang *et al.*[29] developed a *collaborative representation-based classifier* (CRC) by solving the representation vector with the regularized least squares method. Naseem *et al.*[18] devised a linear regression classifier (LRC) by representing a query image as a linear combination of class-specific dictionaries.

Recently, Ram *et al.*[19–22] proposed a smooth ordering method for image processing. The main idea of smooth ordering is to capture the smoothness of the image patches and use it for the recovery of the corrupted image. Specifically, it first orders all patches of a given corrupted image such that they are chained in the "shortest possible path".[22] The reordered patches are expected to be smooth and also induce a 1-D ordering of the image pixels in the center of the patches. Then a 1-D smooth operator (e.g., interpolation) is applied to the reordered pixels, generating the decent recovered image.

However, the original smooth ordering methods only consider the order of the original dataset and change it into a smooth version. They fail to take into account the distance between the high-dimensional data points in the manifold. To reduce this limitation, J. Wang[23,24] developed a *Multiple 1-D embedding* (1DME) method to embed the high-dimensional data into multiple 1-D sequences on a line and perform classification on the 1DME. Due to the specific objects an application deals with, the metric on the data sets should be consistent of the objects. For instance, the original smooth ordering use the Euclidean distance to measure the dissimilarity between image patches. And in the experiments of Ref. **?**, the so-called 1-shift (Euclidean) distance is introduced to measure the dissimilarity of handwritten digits. These distances may be not appropriate for other real-world applications. For example, in this paper, we deal with facial images. The researches have shown that facial images of a subject with distinct expression and illumination can be well represented by a low-dimensional subspace.[7] For such data, samples in a class may not be close in the sense of Euclidean distance but lie in the same subspace. Thus, we need to develop a dissimilarity metric to measure the subspace membership of the points in the facial-image datasets for decent classification performance.

In this paper, we propose a novel classifier based on *Multiple 1-D embedding* (1DME) and *collaborative representation* (CR). The contributions of this paper are as follows:

1. A *collaborative representation based dissimilarity* (CRD) metric is developed to measure the affinity of the points in the facial-image data sets. The experimental results show that under CRD, data samples in a class are similar while those in distinct classes have large dissimilarity value. Compared to the Euclidean distance, it leads to better classification results.
2. We present a CR-based 1DME method (referred as CR1DME) to embed a high dimensional dataset onto multiple 1-D sequences while preserving the proximity

Table 1: Notations used in this paper.

Notation	Description
m	feature dimension
n	number of all samples
l	number of training samples
C	number of classes
$\mathcal{X}_L$	set of labeled samples
$\mathcal{X}_U$	set of unlabeled samples
$\mathcal{I} = \{1, 2, \cdots, n\}$	index set of n samples
$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$	matrix of all samples

of data points. In this way, the classification problem on the original high dimension reduces to a simpler one in the 1DME framework, which can be efficiently tackled with 1-D classifiers. Then we leverage CR1DME to develop a classifier by learning 1DME-based interpolation scheme.

The rest of the paper is arranged as follows. Section 2 formulates the problem and reviews the 1-D manifold embedding (1DME) based method. In Section 3, we propose the CR1DME model and utilize it to develop a novel classifier. In Section 4, we conduct experiments on popular real-world face data sets to validate the efficacy of the proposed method for face recognition. Finally, we give the conclusion in Section 5.

2. Review of Multiple 1-D Embedding-Based Method

In this section, we formulate the classification problem and review the 1DME-based method proposed by J. Wang.[23,24]

To begin with, we specify some important notations used in the paper. Throughout this paper, we represent matrices with boldface uppercase letters, vectors with boldface lowercase letters, and scalars with normal italic letters, respectively. Sets are denoted as uppercase calligraphic letters. For a set $\mathcal{S}$, $|\mathcal{S}|$ denotes its cardinality. Assume the number of classes is C and denote $C = \{1, 2, \cdots, C\}$. Let $\mathcal{X}_L = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l\} \subset \mathbb{R}^m$ be the labeled sample sets, where m denotes the feature dimension and l is the number of labeled samples. Analogously, $\mathcal{X}_U = \{\mathbf{x}_{l+1}, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^m$ denotes the set of unlabeled samples where n is the number of all samples. Denote $\mathcal{X} = \mathcal{X}_L \bigcup \mathcal{X}_U$ and arrange the data points as the columns of a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$. Let $y_i \in C$ be the label of $\mathbf{x}_i$ and denote $\mathcal{Y}_L = \{y_i\}_{i=1}^{l}$. Table 1 summarizes the notations and their descriptions. Given the labeled dataset $(\mathcal{X}_L, \mathcal{Y}_L)$, the goal is to learn a classifier to predict the labels of unlabeled data points $\{\mathbf{x}_i\}_{i=l+1}^{n}$.

With the notations above, we now proceed to introduce the 1DME method. Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^m$ be the given high dimensional dataset. The goal of 1D embedding is to embed it into a line while preserving the important geometric structures of the original dataset, such as the proximity of data points. To this end, we will apply a *smooth ordering* (column) permutation $\mathbf{P}$, which changes the column order of $\mathbf{X}$, to produce a *smooth* data

matrix $\mathbf{X}^p$. Here, the smoothness of a data matrix is measured by the total variation of its columns

$$\|\mathbf{X}\|_{\mathrm{TV}} = \sum_{i=1}^{n-1} d\left(\mathbf{x}_i, \mathbf{x}_{i+1}\right) \tag{2.1}$$

where $d(\mathbf{x}, \mathbf{z})$ denotes the predefined dissimilarity or distance metric between $\mathbf{x}$ and $\mathbf{z}$. Therefore, we compute the permutation operator $\mathbf{P}$ by minimizing the total variation of $\mathbf{P}(\mathbf{X})$:

$$\mathbf{P} = \arg\min_{\mathbf{P}} \|\mathbf{P}(\mathbf{X})\|_{\mathrm{TV}} . \tag{2.2}$$

By the way, a permutation operator $\mathbf{P}$ is induced by a permutation π of the index set $\mathcal{I} = \{1, 2, \cdots, n\}$

$$\mathbf{P}(\mathbf{X}) := \left[\mathbf{x}_{\pi(1)}, \mathbf{x}_{\pi(2)}, \cdots, \mathbf{x}_{\pi(n)}\right].$$

Hence, we may identify $\mathbf{P}$ with π.

As Ref. 22 mentioned, the optimization problem (2.2) is equivalent to searching the shortest path passing through $\mathcal{X}$ and visit each data point only once. Therefore, it turns to be a traveling salesman problem (TSP),[5] which is computationally expensive for large datasets. For this reason, usually the problem (2.2) is approximatively solved by a simple yet effective greedy algorithm.[22]

After obtaining the permutation operator $\mathbf{P}$ and the ordered data matrix $\mathbf{X}^p = [\mathbf{x}_{\pi(1)}, \mathbf{x}_{\pi(2)}, \cdots, \mathbf{x}_{\pi(n)}]$, we make the 1-D embedding of $\mathcal{X}$ as follows: Define a mapping $h : \mathcal{X} \to \mathbb{R}$ such that

$$h\left(\mathbf{x}_{\pi(i)}\right) = t_i, \ i = 1, 2, \cdots, n$$

where $t_1 = 0$ and $t_{j+1} - t_j = d\left(\mathbf{x}_{\pi(j)}, \mathbf{x}_{\pi(j+1)}\right)$ for $j = 1, 2, \cdots, n-1$. Thus, the function h gives a 1-D embedding $\{t_j\}_{j=1}^n$ of $\mathcal{X}$. It can be seen that the vector $\mathbf{t} = [t_1, t_2, \cdots, t_n]$ has the same total variation and smoothness as $\mathbf{X}^p$, i.e., $\|\mathbf{t}\|_{\mathrm{TV}} = \|\mathbf{X}^p\|_{\mathrm{TV}}$.

The algorithm to compute the smooth ordering permutation was first developed in Ref.22. In Ref. 23, a distance measurement on the ordered sequence is added (see Step 3 in **Algorithm 1**). For readers' convenience, we quote it in the following.

Algorithm 1 1-D Manifold Embedding

Input: The data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$, a random probability vector $\mathbf{p} = [p_1, p_2, \cdots, p_n]$, the parameter α and the dissimilarity function d.
Output: A 1-D embedding $\mathbf{t} = [t_1, t_2, \cdots, t_n]$ of $\mathbf{X}$.
Initialization: Randomly choose a index $j \in \mathcal{I} = \{1, 2, \cdots, n\}$ and set $\pi(1) = j$ and $t_1 = 0$.
For $i = 1$ to $n - 1$

1: Set $\mathcal{N}^c(i) = \mathcal{N}\left(\mathbf{x}_{\pi(i)}\right) \setminus \{\pi(1), \pi(2), \cdots, \pi(i)\}$.
2: Compute the cardinality of $\mathcal{N}^c(i)$ denoted as $|\mathcal{N}^c(i)|$, and determine $\pi(i+1)$ using the following criterion.

– If $|\mathcal{N}^c(i)| \geq 2$, we first find the nearest sample $\mathbf{x}_{i_1}$ and the second nearest sample $\mathbf{x}_{i_2}$ of $\mathbf{x}_{\pi(i)}$ in $\mathcal{N}^c(i)$. Then we compute q_i as

$$q_i = \left(1 + \exp\left(\frac{d\left(\mathbf{x}_{\pi(i)}, \mathbf{x}_{i_1}\right) - d\left(\mathbf{x}_{\pi(i)}, \mathbf{x}_{i_2}\right)}{\alpha m}\right)\right)^{-1}.$$

If $q_i \leq p_{\pi(i)}$, we set $\pi(i+1) = i_2$. Otherwise, we set $\pi(i+1) = i_1$.

– If $|\mathcal{N}^c(i)| = 1$, set $\pi(i+1)$ as the index of the only sample in $\mathcal{N}^c(i)$.

– If $|\mathcal{N}^c(i)| = 0$, we define $\pi(i+1)$ using the points in $\mathcal{I}^c(i) := \mathcal{I} \setminus \{\pi(1), \pi(2), \cdots, \pi(i)\}$. If $|\mathcal{I}^c| \leq 2$, we find the nearest and the second nearest samples of $\mathbf{x}_{\pi(i)}$ in $\mathcal{I}^c$ and define $\pi(i+1)$ as described in the case $|\mathcal{N}^c(i)| \geq 2$. Otherwise, $|\mathcal{I}^c| = 1$ and $i = n-1$. Then we set $\pi(i+1)$ as the index of the only sample in $\mathcal{X}^c$.

3: Set $t_{i+1} = t_i + d\left(\mathbf{x}_{\pi(j)}, \mathbf{x}_{\pi(j+1)}\right)$.

For end

A multiple 1D embedding contains several versions of 1D embedding with the different heads, which are chosen at random. To develop a 1DME-based algorithm, we first construct an individual classifier based on each 1D embedding, then integrate these classifiers into a single one by a voting role. The process above is repeated for successively improving the quality of the classifier. The motivation and benefits of 1DME-based algorithm are as follows. The classification problem for high-dimensional data reduces to that cab be solves in a 1-D framework. Hence, the developed algorithms are more stable and controllable. A M1ED-based classifier is essentially dimension-free. Hence, it can alleviate the curse of dimensionality,[13] which appears as the data dimensionality increases. More details can be found in Ref. 24.

3. Proposed Method

This section aims to describe the proposed classifier for face recognition, which is arranged as follows. First, we develop a collaborative representation (CR) based approach to learn the dissimilarity between pairwise data points. This leads to a collaborative representation (CR) based 1DME (CR1DME) method by incorporating the proposed CR based dissimilarity method into the original 1DME-based one.[23] Finally, we propose a classifier of CR1DME with application to face recognition.

3.1. *Dissimilarity Learning via Collaborative Representation*

In 1DME-based method, the dissimilarity metric $d(\cdot, \cdot)$ plays an important role in learning a decent 1DME of the original data set. Denote by $\mathbf{D} = (d_{ij}) \in \mathbb{R}^{n \times n}$ the dissimilarity matrix where $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$. We aim to propose a dissimilarity metric under which samples in a class are similar while those in distinct classes have large dissimilarity value.

A conventional choice of the dissimilarity metric d is the Euclidean distance due to its simplicity. In this case, if two data points have small Euclidean distance, their positions in

a 1-D embedding will be close. Then the two points will be assigned to the same class with large probability. However, in many real-world applications, Euclidean distance is not a good measurement for data dissimilarity. For example, facial images of a subject under distinct expression and illumination can be well represented by a low-dimensional subspace.[1] For such data, samples in a class may not be close in the sense of Euclidean distance but lie in the same subspace. This motivates us to propose a novel dissimilarity metric which measures the subspace membership of the data points instead of the Euclidean distance. For this end, we first compute the coefficient matrix $\mathbf{C}$ using collaborative representation

$$\min_{\mathbf{C}\in\mathbb{R}^{m\times n}} \|\mathbf{X}-\mathbf{X}\mathbf{C}\|_F^2 + \lambda\|\mathbf{C}\|_F^2 \tag{3.1}$$

where λ denotes the regularization parameter. The optimal solution of the problem (3.1) can be explicitly formulated as

$$\mathbf{C}^* = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{X}$$

where $\mathbf{I} \in \mathbb{R}^{n\times n}$ denotes the identity matrix. Thus, the problem (3.1) can be efficiently computed. For symmetry and normalization, we construct the affinity matrix $\mathbf{A} = \left(|\mathbf{C}^*| + |\mathbf{C}^*|^T\right)$ and normalize $\mathbf{A}$ to make the maximal and minimal value of its entries as 1 and 0, respectively. The proposed dissimilarity is defined as $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - a_{ij}$, where a_{ij} denotes the (i, j)-th entry of $\mathbf{A}$. According to the theory in Ref. 17, when under appropriate conditions, $d(\mathbf{x}_i, \mathbf{x}_j)$ is large when $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to distinct subspaces, and small if they are from a same subspace. Therefore, the proposed dissimilarity encourages the 1D embedded data from a same subspace (corresponding to the same class) to be close while those from distinct subspaces far apart. This gives rise to success of the proposed CR1DME method.

3.2. *Classifier of CR1DME*

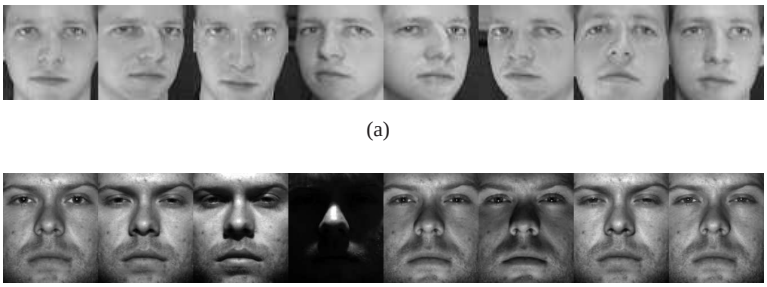
Now we develop a classifier of CR1DME by integrating the individual modulus aforementioned and a M1dE-based adaptive interpolation scheme[23] into a unified framework.

Given the data matrix $\mathbf{X}$, we first learn a dissimilarity matrix $\mathbf{D} = (d_{ij}) \in \mathbb{R}^{n\times n}$ with $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$ using collaborative representation as stated in the previous section. Then, with the dissimilarity defined, we compute a classifier using a 1DME-based classification algorithm, e.g., the 1DMEI algorithm in Ref. **?** without the iterative procedure.

Algorithm 2 Classifier of CR1DME

Input: The data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{m\times n}$, labels $\{y_i\}_{i=1}^{l}$ of training samples , and the number of embeddings K.
Output: The predicted labels of test samples

1: Learn the dissimilarity matrix $\mathbf{D}$ via collaborative representation.
2: Construct K 1DMEs of $\mathbf{X}$ using Algorithm 1.
3: Build K functions f_k $(1 \le k \le K)$ defined on $\mathcal{X}$ using Eqs. (3.2)-(3.4).
4: Predict the labels of test samples $\mathbf{x}_i$ $(l+1 \le i \le n)$ using the sign of $f^*(\mathbf{x}_i) = \frac{1}{K}\sum_{k=1}^{K} f_k(\mathbf{x}_i)$.



(a)

(b)

Fig. 1: (a) Cropped images of one subject in the ORL database; (b) cropped images of one subject in the Extend Yale B database.

To have a better understanding of our method, we formulate the classification procedure as follows. It is known that a problem of multiclass classification can be reduced to multiple binary classification problems using one-vs.-rest (OvR) strategy.[12] For this reason and brevity of notations, we describe the proposed classifier for a binary problem only, where the label $y \in \{-1, +1\}$. This algorithm can be found in Ref. 23 too. We define a univariate function $\tilde{f} : \mathbb{R} \to \mathbb{R}$ such that

$$\tilde{f}\left(t_{\pi(i)}\right) = y_i,\ 1 \le i \le l. \tag{3.2}$$

Then a cubic interpolation operator $\mathbf{L}$ is utilized to interpolate $\tilde{f}$ on the set $\{t_{\pi(1)}, t_{\pi(2)}, \cdots, t_{\pi(l)}\}$

$$\left(\mathbf{L}\tilde{f}\right)\left(t_{\pi(i)}\right) = \tilde{f}\left(t_{\pi(i)}\right). \tag{3.3}$$

This yields a function $f : \mathcal{X} \to \mathbb{R}$ formulated as

$$f = \left(\mathbf{L}\tilde{f}\right) \circ (h \circ \mathbf{P}). \tag{3.4}$$

Finally, the label of the test sample $\mathbf{x}_i$ $(l+1 \le i \le n)$ is determined by the sign of $f(\mathbf{x}_i)$. For better performance, we can employ K permutation operators $\mathbf{P}_k$ and compute K 1DMEs. Then we can obtain K functions f_k $(1 \le k \le K)$ defined on $\mathcal{X}$ using the same strategy aforementioned. We use the sign of $f^*(\mathbf{x}_i) = \frac{1}{K}\sum_{k=1}^{K} f_k(\mathbf{x}_i)$ to determine the label of $\mathbf{x}_i$. Algorithm 2 provides a pseudocode of the proposed classifier of CR1DME.

4. Experiments

In this section, we conduct experiments on popular read-world datasets to evaluate the efficacy of the proposed classifier for face recognition. We first compare the proposed dissimilarity matrix with the Euclidean distance for 1DME. Then we compare the performance of the proposed method with several state-of-the-art classifiers for face recognition.

4.1. *Experimental Settings*

This section aims to introduce the used datasets, competing methods and their parameter settings in the experiments. We utilize two popular databases, i.e., the ORL database[25] and the Extended Yale B database.[16] Fig. 1 shows some facial images from these two databases. The description of them are as follows.

1. ORL Database: This database contains facial images of 40 subjects taken under varying expression and illumination. For each subject, there are 10 facial images. Each image is resized to have 32×32 pixels.
2. Extended Yale B Database: The Extended Yale B database contains 38 subjects and around 64 frontal-face images per subject captured under distinct lighting conditions. Thus the database consists of 2432 facial images. The images are cropped with 48×42 pixels.

To evaluate the performance of CR1DME, we compare it with several state-of-the-art classifiers for face recognition. They are LRC, CRC, SRC, and two popular semi-supervised classifiers, i.e., laplacian regularized least squares (LapRLS) and laplacian support vector machines (LapSVM).[3] For each method, we use the code provided by the corresponding authors. For CRC and SRC, the regularization parameters are chosen as 0.001 and 0.1, respectively. For LapRLS and LapSVM, we choose $N = 10$ nearest neighbor, and two regularization parameters $\gamma_A = 10^{-6}$ and $\gamma_I = 10^{-2}$ according to the reference.[3] For CR1DME, we choose the parameter $\alpha = 10$, $\lambda = 4 \times 10^{-3}$ and the number of embeddings $K = 15$.

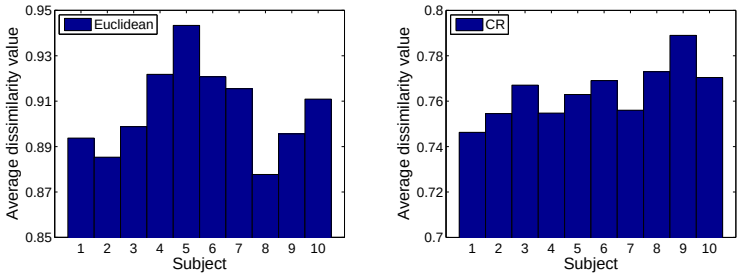
4.2. *Dissimilarity Comparison for 1DME*

Now we analyze the impact of the dissimilarity metric to the classifier of 1DME. Specifically, we compare the proposed *CR based dissimilarity* (CRD) with the Euclidean distance for 1DME. Denote by $\mathcal{X}_c$ the set of samples belonging to class c ($1 \le c \le C$) in $\mathcal{X}$. For comparison, we define an *average class dissimilarity* between a sample $\mathbf{x}_i$ and $\mathcal{X}_c$

$$\bar{d}(\mathbf{x}_i, \mathcal{X}_c) := \frac{1}{|\mathcal{X}_c|} \sum_{\mathbf{x} \in \mathcal{X}_c} d(\mathbf{x}_i, \mathbf{x}) \tag{4.1}$$

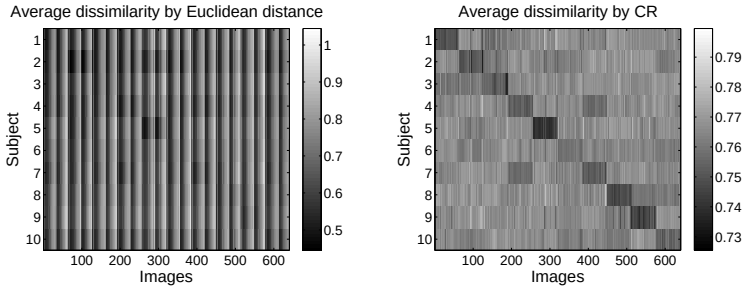
for $i = 1, 2, \cdots, n$ and $c = 1, 2, \cdots, C$. Ideally, we expect that $\bar{d}(\mathbf{x}_i, \mathcal{X}_c)$ has small value if $\mathbf{x}_i$ belongs to class c and large value otherwise. If so, the embeddings of the samples in a same class will be close and those of samples in distinct class will be far apart, thereby leading to decent classification results. Consequently, $\bar{d}$ can be used as an evaluation index for the dissimilarity metric.

We use a subset of the facial images of the first ten subjects in the Extended Yale B database to conduct the experiment. Thus, this subset consists of 640 facial images of ten subjects. For efficiency, we downsample each facial image to have 6×5 pixels with the downsample ratio $1/8$. We compute the average class dissimilarity between the 4-th facial image of the first subject $\mathbf{x}_4$ and $\mathcal{X}_c$ for $c = 1, 2, \cdots, C$ using the Euclidean distance and CRD, respectively. The results are shown in Fig. 2. It can be seen that with the Euclidean



(a) (b)

Fig. 2: The average class dissimilarity of an image from the first subject computed using (a) the Euclidean distance; (b) the proposed CR based dissimilarity.



(a) (b)

Fig. 3: The average class dissimilarity map of a subset of the Extended Yale B database computed using (a) the Euclidean distance; (b) the proposed CR based dissimilarity.

distance based dissimilarity, $\mathbf{x}_4$ has the smallest average class dissimilarity with the images of the 8-th subject rather than those of the first subject. In contrast, using the proposed dissimilarity, $\mathbf{x}_4$ is the most similar to the facial images of the first subject, which is also the true subject of $\mathbf{x}_4$. By applying 1DME on this subset, the 1-D embedding of $\mathbf{x}_4$ will be close to those from the same class with large probability.

To make the results more convincing, we also compute the average class dissimilarity between each sample and samples from every class. Fig. 3 shows the average class dissimilarity map using the Euclidean distance and the CRD, respectively. It can be found that the CRD performs better than the Euclidean distance. Concretely, using the CRD, the samples always have the smallest average class dissimilarity with those from the same class. In comparison, the map computed with the Euclidean distance is disordered.

For simplicity, we refer to the classifier of 1DME using the Euclidean distance as E1DME. We conduct face recognition experiments on this subset using E1DME and CR1DME, respectively. For each subject, n_t $(= 6, 8, 10, 12)$ images per class are randomly selected for training and the rest for test. Fig. 4 displays the recognition results on this

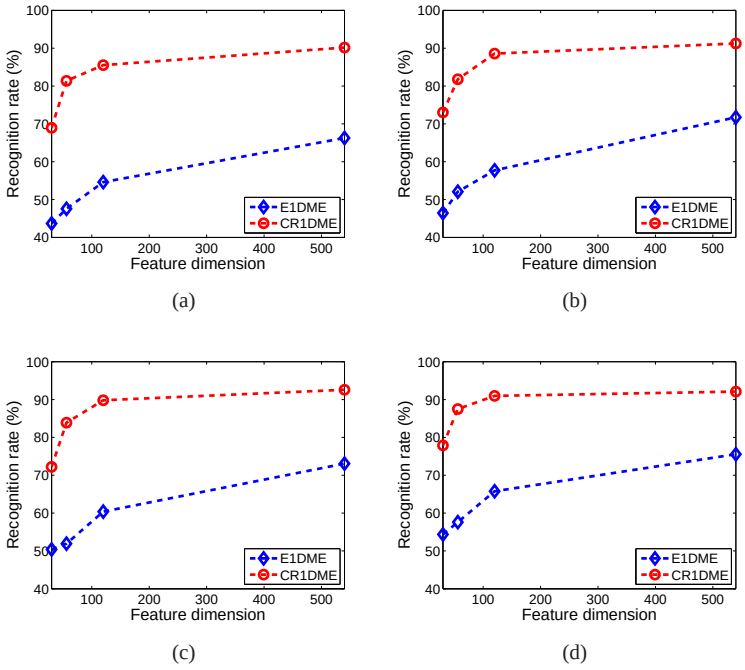


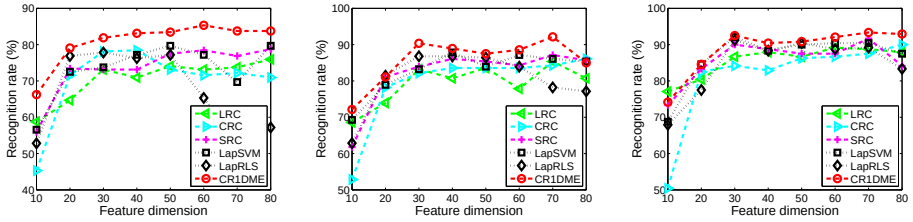
Fig. 4: Recognition results on a subset of the Extended Yale B database with the first 10 subjects for E1DME and CR1DME based on n_t images per subject for training. (a) $n_t = 6$; (b) $n_t = 8$; (c) $n_t = 10$; (d) $n_t = 12$.

subset. It can be seen that CR1DME performs much better than E1DME at all levels.

4.3. *Results on the ORL Database*

In this section, we validate the effectiveness of the proposed method on the ORL database. For each subject, we randomly select n_t $(= 2, 3, 4, 5)$ images for training and the rest for test. Before performing classification, we use PCA to reduce the dimension of vectorized images to vary from 10 to 80. Fig. 5 shows the recognition rate of distinct algorithms as a function of the feature dimension for distinct number of images per subject for training. To make it more convincing, we repeat the face recognition experiments of each method ten times with $n_t = 2$ and feature dimension varying from 10 to 80. Table 2 reports the average recognition rates of each method over the ten-run test. Based on the results, we draw the following conclusions.

First, in general as the training set size and feature dimension increases, all methods achieve higher recognition rates. However, CR1DME outperforms competing methods in most cases. As shown in Table 2, when the feature dimension is 80, the average recognition rate of CR1DME is 84.59 percent, while all the other methods achieve the average recognition rate less than 77 percent.



(a) (b) (c)

Fig. 5: Recognition results on the ORL database for various algorithms based on n_t images per subject for training. (a) $n_t = 2$; (b) $n_t = 3$; (c) $n_t = 4$.

Table 2: Recognition rate (%) on the ORL database for varying feature dimension with 2 images per subject for training.

Feature dimension	10	20	30	40	50	60	70	80
LRC	57.28	69.34	71.91	72.03	72.19	72.97	72.22	70.44
CRC	48.44	70.78	74.38	75.65	74.97	72.78	72.22	72.91
SRC	61.06	73.38	76.50	76.72	77.81	78.59	77.13	76.66
LapSVM	59.09	73.09	72.78	75.03	77.06	77.75	78.66	75.84
LapRLS	56.63	73.69	75.88	74.31	71.41	65.25	58.59	52.21
CR1DME	**63.34**	**77.84**	**80.44**	**82.34**	**84.75**	**82.72**	**85.16**	**84.59**

Second, since both CR1DME and CRC take advantage of the collaborative representation, the comparison between them suggests that the 1DME based classifier performs better than the class residual based decision rule in CRC. Concretely, for $n_t = 2$, when the feature dimension varies from 10 to 80, the average recognition rate of CR1DME improves that of CRC by at least 5 percent at each level.

4.4. *Results on the Extended Yale B Database*

This section aims at testing the performance of CR1DME on the Extended Yale B database. We compute the recognition rates with feature dimensions 30, 56, 120 and 504. The corresponding downsampling ratio is $1/8$, $1/6$, $1/4$ and $1/2$, respectively. Fig. 6 shows the recognition rates of distinct algorithms under varying feature dimensions and values of n_t. We also repeat the experiments above for $n_t = 6$ and $n_t = 8$ ten times and report the average recognition rates in Table 3. From the results, some conclusions can be obtained as follows.

First, CR1DME significantly outperforms competing methods for the training set with small size. For example, in Fig. 6(a) with the number of images per subject $n_t = 6$ and feature dimension 30, the recognition rate of CR1DME is 71% while recognition rates of all other methods are less than 55%. In addition, as the feature dimension increases, CR1DME

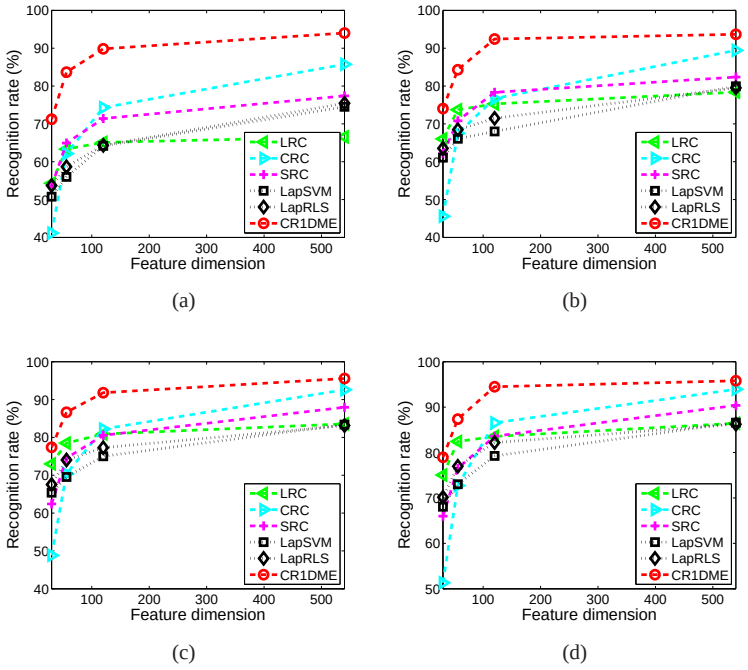


Fig. 6: Recognition results on the Extended Yale B database for various algorithms based on n_t images per subject for training. (a) $n_t = 6$; (b) $n_t = 8$; (c) $n_t = 10$; (d) $n_t = 12$.

Table 3: Recognition rate (%) on the Extended Yale B database for varying feature dimension.

	$n_t = 6$				$n_t = 8$			
Feature dimension	30	56	120	504	30	56	120	504
LRC	56.89	60.83	63.62	67.34	67.03	71.63	74.53	76.34
CRC	42.92	62.23	74.25	84.37	47.11	67.73	78.47	88.63
SRC	53.78	65.17	71.49	79.96	59.94	70.38	77.09	85.06
LapSVM	46.60	57.05	64.41	74.98	50.51	62.13	70.19	80.54
LapRLS	48.66	59.87	64.84	74.34	53.85	66.03	71.96	79.45
CR1DME	**70.34**	**83.16**	**90.40**	**94.05**	**73.89**	**85.09**	**92.12**	**94.39**

still achieves better performance than other algorithms. Since images for training are often limited in practice, this makes senses for real-world face recognition.

Second, when the feature dimension increases from a low level, the recognition rate of CR1DME increases rapidly. This comes from the fact that for the images with higher dimension, CR1DME can learn a much better dissimilarity matrix. Then the 1DME of the images in the same class are closer and those from different classes are far apart, thereby

producing better classification results.

5. Conclusion

In this paper, we propose a novel method for face recognition based on CR and 1DME.[23] First, we devised a CR based dissimilarity metric by considering the subspace structure of facial images. Under the metric, data samples in a same class are similar while those in distinct classes have large value of dissimilarity. With the dissimilarity metric, we make a multiple 1D-embedding of the original datasets, preserving the proximity of data points in the high dimensional ambient space. Then the classification problem on high dimension can be solved in the one-dimensional framework. A classifier of CR1DME is developed based on 1DME interpolation scheme that was developed by one of the authors in Ref. 23. Experimental results on real datasets show that the proposed method outperforms several other state-of-the-art methods for face recognition, especially when labeled samples per class are limited.

References

1. R. Basri and D. Jacobs, Lambertian reflection and linear subspaces, *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(3) (2003) 218–233.
2. P. Belhumeur, J. Hespanha, D. Kriegman and A. Rosenfeld, Eigenfaces versus fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7) (1997) 711–720.
3. M. Belkin, P. Niyogi and V. Sindhwani, Manifold regularization: a geometric framework for learning from labeled and unlabeled examples, *J. Mach. Learn. Res.* **7** (2006) 2399–2434.
4. R. Brunelli and T. Poggio, Face recognition: Features versus templates, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(10) (1993) 1042–1052.
5. T. Cormen, *Introduction to Algorithms* (Cambridge, MA, USA: MIT Press, 2001).
6. E. Elhamifar and R. Vidal, Block-sparse recovery via convex optimization, *IEEE Trans. Signal Process.* **60**(8) (2012) 4094–4107.
7. E. Elhamifar and R. Vidal, Sparse subspace clustering: algorithm, theory, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(11) (2013) 2765–2781.
8. B. Gunturk, A. Batur, Y. Altunbasak, M. Hayes and R. Mersereau, Eigenface-domain super-resolution for face recognition, *IEEE Trans. Image Process.* **12**(5) (2003) 597–606.
9. R. He, W. Zheng and B. Hu, Maximum correntropy criterion for robust face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8) (2011) 1561–1576.
10. X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang, Face recognition using laplacianfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3) (2005) 328–340.
11. H. Ho and R. Chellappa, Pose-invariant face recognition using markov random fields, *IEEE Trans. Image Process.* **22**(4) (2013) 1573–1584.
12. C. Hsu and C. Lin, A comparison of methods for multiclass support vector machines, *IEEE Trans. Neural Netw.* **13**(2) (2002) 415–425.
13. G. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(1) (1968) 55–63.
14. J. Kim, J. Choi, J. Yi and M. Turk, Effective representation using ica for face recognition robust to local distortion and partial occlusion, *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(12) (2005) 1977–1981.
15. T. Kim, J. Kittler and R. Cipolla, On-line learning of mutually orthogonal subspaces for face recognition by image sets, *IEEE Trans. Image Process.* **19**(4) (2010) 1067–1074.

16. K. Lee, J. Ho and D. Driegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Anal. Mach. Intell.* **27** (May 2005) 684–698.
17. C. Lu, H. Min, Z. Zhao, L. Zhu, D. Huang and S. Yan, Robust and efficient subspace segmentation via least squares regression, *Proc. European Conf. Comput. Vis.*, 2012, pp. 1801–1808.
18. I. Naseem, R. Togneri and M. Bennamoun, Linear regression for face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(11) (2010) 2106–2112.
19. I. Ram, I. Cohen and M. Elad, Patch-ordering-based wavelet frame and its use in inverse problems, *IEEE Trans. Image Process.* **23**(7) (2014) 2779–2792.
20. I. Ram, M. Elad and I. Cohen, Generalized tree-based wavelet transform, *IEEE Trans. Signal Process.* **59**(9) (2011) 4199–4209.
21. I. Ram, M. Elad and I. Cohen, Redundant wavelets on graphs and high dimensional data clouds, *IEEE Signal Process. Lett.* **19**(5) (2012) 291–294.
22. I. Ram, M. Elad and I. Cohen, Image processing using smooth ordering of its patches, *IEEE Trans. Image Process.* **22**(7) (2013) 2764–2774.
23. J. Wang, Semi-Supervised Learning Using Multiple One-Dimensional Embedding-Based Adaptive Interpolation, **in this special issue of** *International Journal of Wavelets, Multiresolution, and Information Processing*, 2015.
24. J. Wang, Semi-Supervised Learning Using Ensembles of Multiple 1D-Embedding-Bases Label Boosting, **in this special issue of** *International Journal of Wavelets, Multiresolution, and Information Processing*, 2015.
25. F. Samaria and A. Harter, Parameterisation of a stochastic model for human face identification, *Proc. IEEE Workshop Applicat. Comput. Vis.*, 1994, pp. 138–142.
26. M. Turk and A. Pentland, Face recognition using eigenfaces, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1991, pp. 586–591.
27. J. Wright, A.Yang, A. Ganesh, S. Sastry and Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(2) (2009) 210–227.
28. A. Yang, Z. Zhou, A. Balasubramanian, S. Sastry and Y. Ma, Fast ℓ_1-minimization algorithms for robust face recognition, *IEEE Trans. Image Process.* **22**(8) (2013) 3234–3246.
29. L. Zhang, M. Yang and X. Feng, Sparse representation or collaborative representation: which helps face recognition?, *Proc. IEEE Conf. Int. Conf. Comput. Vis.*, 2011, pp. 471–478.