We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

4

Face Recognition by Discriminative Orthogonal Rank-one Tensor Decomposition

Gang Hua
Microsoft Live Labs Research,
One Microsoft Way, Redmond, WA 98052,
U.S.A.

1. Introduction

Discriminative subspace analysis has been a popular approach to face recognition. Most of the previous work such as Eigen-faces (Turk & Pentlend, 1991), LDA (Belhumeur et al., 1997), Laplacian faces (He et al., 2005a), as well as a variety of tensor based subspace analysis methods (He et al., 2005b; Chen et al., 2005; Xu et al., 2006; Hua et al., 2007), can all be unified in the graph embedding framework (Yan et al., 2007). In this Chapter, we investigate the effects of two types of regularizations on discriminative subspace based face recognition techniques: a new 2D tensor representation for face image, and an orthogonal constraint on the discriminative tensor projections.

Given a face image, the new tensor representation firstly divides it into non-overlapping blocks. Then following the raster-scan order, the raster-scanned pixel vectors of each of the image blocks are put into the columns of a new 2D tensor. It is easy to figure out that the row vectors of the new 2D tensor are in essence different down-sampled images of the original face images. Pursuing discriminative 2D tensor projections with the new tensor representation is of special interest, because the left projection indeed functions as local filters in the original face image and the right projection reveals to us that which local block is more important for recognition.

This new representation puts concrete physical meanings to the left and right projections of the discriminative tensor projections. While the 2D tensor representation using the original images does not present any meaningful physical explanations on column and row pixel vectors. We call this new tensor representation Global-Local representation (Chen et al., 2005; Hua et al., 2007).

On the other hand, we reveal a very important property regarding the orthogonality between two tensor projections, and thus present a novel discriminative orthogonal tensor decomposition method for face recognition. To the best of our knowledge, this method, firstly introduced in (Hua et at., 2007), is the first discriminative orthogonal tensor decomposition method ever proposed in the literature.

Both of the two regularization techniques put additional constraints on the capacity (a.k.a., the VC-dimension) of the discriminative projections and thereby improve the generalization ability of the learned projections. We perform empirical analysis and comparative study on widely adopted face recognition bench-mark such as Yale, ORL, YaleB and PIE databased to

better understand the behaviours of the two. Note most of our results are adopted from (Hua et al., 2007) but we provide more analysis and discussions in this Chapter.

The rest of the Chapter is organized as follows: Section 2 defines some terminologies and mathematic notations on tensor analysis, as well as a very important property of orthogonal tensor projections, which will be used across the Chapter. Section 3 reviews the Global-Local tensor representation with its benefits discussed. Then, in Section 4, we present the new method for discriminative orthogonal rank-one tensor decomposition. Section 5 will discuss the experimental results on bench-mark face databases. Section 6 highlights some general remarks regarding the orthogonal rank-one tensor decomposition method for the task of face recognition. We conclude this Chapter in Section 7.

2. Introduction to tensor analysis

In multi-dimensional linear algebra, a tensor of order n or a nD tensor is a multiple dimensional array $\mathbf{X} \in \mathbf{R}^{n_0 \times n_1 \times \ldots \times n_n}$. We denote the element at position $(i_1, i_2, \ldots, i_n)$ to be $x_{i_1 i_2 \ldots i_n}$. For example, a matrix is a tensor of order 2 or 2D tensor, and x_{ij} denotes its element at the ith row and j^{th} column. In the following we introduce several definitions in tensor analysis, which is essential to present the discriminative orthogonal tensor decomposition method. Similar definitions are also adopted in (Hua et al., 2007).

The first definition we introduce here is the concept of k-mode product for a tensor and a matrix (a.k.a, an order 2 tensor). Following the tensor algebra literature (Kolda, 2001), we have:

Definition 1: The k-mode product of a tensor $\mathbf{X} \in \mathbf{R}^{n_1 \times n_2 \times \ldots \times n_k \times \ldots \times n_n}$ and a matrix $\mathbf{B} \in \mathbf{R}^{n_k \times m_k}$ is a $\mathbf{X} \in \mathbf{R}^{n_1 \times n_2 \times \ldots \times n_k \times \ldots \times n_n} \to \mathbf{Y} \in \mathbf{R}^{n_1 \times n_2 \times \ldots \times m_k \times \ldots \times n_n}$ mapping, such that

$$y_{i_1 i_2 \ldots i_{k-1} i'_k i_{k+1} \ldots i_n} = \sum_{j=1}^{n_k} x_{i_1 i_2 \ldots i_{k-1} j i_{k+1} \ldots i_n} b_{j i'_k}. \tag{1}$$

The k-mode product is generally denoted as $\mathbf{Y} = \mathbf{X} \times_k \mathbf{B}$.

The second definition we introduce here is the rank-one tensor. In general, a tensor is said to be of rank one, if it can be decomposed as the tensor product of a set of vectors.

Definition 2: A tensor $\mathbf{X} \in \mathbf{R}^{n_1 \times n_2 \times \ldots \times n_n}$ of order n is said to be with rank one, if and only if there exists a vector set $\hat{X} = \{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots \ldots, \hat{\mathbf{x}}_n\}$ where each $\hat{\mathbf{x}}_i$ is a vector of dimension n_i, and its jth element is denoted as $\hat{x}_{ij}$, such that

$$x_{i_1 i_2 \ldots i_n} = \prod_{j=1}^{n} \hat{x}_{j i_j}. \tag{2}$$

The tensor $\mathbf{X}$ is called the reconstruction rank one tensor of $\hat{X}$, and $\hat{X}$ is said to be the reconstruction vector set.

Based on the definitions above, we introduce the definition of rank one tensor projection:

Definition 3: Given an order n tensor $\mathbf{X}$, a rank one projection is an $\mathbf{X} \in \mathbf{R}^{n_1 \times n_2 \times \ldots \times n_n} \to y \in \mathbf{R}$ mapping, which is defined by a projection vector set $\hat{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots \ldots, \mathbf{p}_n\}$ where each $\mathbf{p}_i$ is a column vector of dimension n_i. Let p_{ij} be the jth element of the vector $\mathbf{p}_i$, we have

$$y = \sum_{i_1, i_2, \ldots i_n} x_{i_1 i_2 \ldots i_n} \times p_{1 i_1} \times p_{2 i_2} \times \ldots \times p_{n i_n} \tag{3}$$

Let $\mathbf{P} \in \mathbf{R}^{n_1 \times n_2 \times \ldots \times n_n}$ be the reconstruction rank one tensor of $\hat{P}$, we have

$$y = \sum_{i_1, i_2, \ldots i_n} x_{i_1 i_2 \ldots i_n} \times p_{i_1 i_2 \ldots i_n}. \tag{4}$$

For ease of presentation, we denote the rank one projection using $\circledast$, i.e., $y = \hat{P} \circledast \mathbf{X}$ or $y = \mathbf{P} \circledast \mathbf{X}$. Obviously, using the k-mode product notation, if we treat each $\mathbf{p}_i$ as a $n_i \times 1$ matrix, we also have

$$y = \mathbf{X} \times_1 \mathbf{p}_1 \times_2 \mathbf{p}_2 \times_3 \dots \times_n \mathbf{p}_n. \tag{5}$$

Indeed, a rank one tensor projection can be deemed as a constrained linear projection. To understand it, we introduce the definition of *unfolding vector*.

Definition 4: The unfolding vector of an order n tensor $\mathbf{X} \in \mathbf{R}^{n_1 \times n_2 \times \dots \times n_n}$ is a vector $\tilde{\mathbf{x}} \in \mathbf{R}^k$, where $k = n_1 n_2 \dots n_n$, such that $\tilde{x}_i = x_{i_1 i_2 \dots i_n}$, where $i_j = \left\lfloor \frac{i - \sum_{k=1}^{j-1} [i_k \prod_{l=k+1}^{n} n_l]}{\prod_{l=j+1}^{n} n_l} \right\rfloor$ can be obtained recursively for $j = 1 \dots n$. Note that here $\lfloor a \rfloor$ means the largest integer that is not larger than a.

Given the vector set representation $\hat{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots \dots, \mathbf{p}_n\}$ of a rank-one tensor projection $\mathbf{P} \in \mathbf{R}^{n_1 \times n_2 \times \dots \times n_n}$, it is easy to figure out that the unfolding vector $\tilde{\mathbf{p}}$ can be obtained by

$$\tilde{\mathbf{p}} = \mathbf{p}_n \otimes \mathbf{p}_{n-1} \otimes \dots \otimes \mathbf{p}_1, \tag{6}$$

where $\otimes$ is the matrix Kronecker product. It is straightforward to figure out the following properties for rank one tensor projection, i.e.,

$$\hat{P} \circledast \mathbf{X} = \tilde{\mathbf{p}}^T \tilde{\mathbf{x}}. \tag{7}$$

It is because of this equivalence that a rank-one tensor projection can be regarded as a parameter constrained vector space linear projection. With the concept of unfolding vector, we finally define orthogonal rank-one tensor projections.

Definition 5: Two rank-one tensor projections $\hat{P}$ and $\hat{Q}$ are said to be orthogonal if and only if their corresponding unfolding vectors $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$ are orthogonal to each other. Mathematically, we have

$$\hat{P} \perp \hat{Q} \Leftrightarrow \tilde{\mathbf{p}} \perp \tilde{\mathbf{q}} \tag{8}$$

This definition essentially relates orthogonal rank-one tensor projections with orthogonal vector projections. Note Definition 5 of orthogonal rank-one tensor projection is equivalent to the definition of orthogonal rank-one tensors in (Kolda, 2001).

We end the section by presenting a sufficient and necessary condition for orthogonal rank-one tensor projections, along with its proof (Hua et al., 2007).

Theorem 1: Given two rank-one tensor projections $\hat{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots \dots, \mathbf{p}_n\}$ and $\hat{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \dots \dots, \mathbf{q}_n\}$, where $\mathbf{p}_i$ and $\mathbf{q}_i$ have the same dimensionality n_i, they are orthogonal if and only if $\mathbf{p}_i \perp \mathbf{q}_i$ held at least for one of the dimension i. Or in short, we have

$$\hat{P} \perp \hat{Q} \Leftrightarrow \exists i, \text{such that } \mathbf{p}_i \perp \mathbf{q}_i$$

Proof: Let $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$ be the unfolding vectors of $\hat{P}$ and $\hat{Q}$, it is easy to figure out that $\tilde{\mathbf{p}}^T \tilde{\mathbf{q}} = \prod_{i=1}^{n} \mathbf{p}_i^T \mathbf{q}_i$ based on the property of Kronecker product (See Definition 4).

"$\Rightarrow$": if $\hat{P} \perp \hat{Q}$, by Definition 5, we have $\tilde{\mathbf{p}}^T \tilde{\mathbf{q}} = \prod_{i=1}^{n} \mathbf{p}_i^T \mathbf{q}_i = 0$. If there does not exists an i, such that $\mathbf{p}_i \perp \mathbf{q}_i$, we would have $\mathbf{p}_i^T \mathbf{q}_i \neq 0$ for all i. Then we would have $\prod_{i=1}^{n} \mathbf{p}_i^T \mathbf{q}_i \neq 0$, which is conflicting with the setting. Therefore, there exists at least one i, such that $\mathbf{p}_i^T \mathbf{q}_i = 0$, i.e., $\mathbf{p}_i \perp \mathbf{q}_i$.

"$\Leftarrow$": If there exists one i , such that $\mathbf{p}_i \perp \mathbf{q}_i$, we have $\mathbf{p}_i^T \mathbf{q}_i = 0$. Then we immediately have $\prod_{i=1}^{n} \mathbf{p}_i^T \mathbf{q}_i = 0$. That essentially means that $\widetilde{\mathbf{p}}^T \widetilde{\mathbf{q}} = 0$, and thus $\hat{P} \perp \hat{Q}$.∎

Theorem 1 reveals that for a pair of rank-one tensors to be orthogonal, it is suffice that the two corresponding vectors in one dimension of their reconstruction vector sets to be orthogonal.

3. Global-local tensor representation

Earlier subspace based methods for face recognition normally treat a face image as a vector data, which completely ignores the spatial structure of the 2 dimensional face image. It is until recently that tensor based representation for face images has become popular (He et al., 2005b; Chen et al., 2005; Xu et al., 2006; Hua et al., 2007). In tensor based representation, a face image is either regarded as an order 2 tensor (raw image) or an order 3 tensor (multi-band filter responses).

With the tensor representation, multi-linear (e.g., bilinear for order 2 tensors) are pursued for discriminative subspace analysis. Tensor based representation enjoys several advantages over vector based representation. First, it has the potential to utilize the spatial structure of the face images. Second, it suffers less from the curse-of-dimensionality because the multi-linear projection has much less parameters to estimate than normal linear vector projections. To give a concrete example, for face images of size 32×32, pursuing one discriminative projection for vector based representation needs to estimate $32 \times 32 = 1024$ parameters. While for order 2 tensor representation (raw image), pursuing one bilinear projection only needs to estimate $32 + 32 = 64$ parameters. Thirdly, because multi-linear projection has much less parameters to estimate, it is less likely to over-fit with the training data, especially when we only have small number of training examples.

Nevertheless, the majority of the previous works regard the raw face image as the order 2 tensor. Given a order 2 tensor $\mathbf{X} \in \mathbf{R}^{n_1 \times n_2}$, the rank-one tensor projection $\hat{P} = \{\mathbf{p}_1, \mathbf{p}_2\}$ is also called a bilinear projection such that $y = \mathbf{p}_1^T \mathbf{X} \mathbf{p}_2$, where $\mathbf{p}_1$ and $\mathbf{p}_2$ are named the left projection and right projection, respectively. Essentially the left and right projections of the bi-linear projection are performing analysis on the column pixel space and raw pixel space of the raw images, respectively. It does not really explore much of the spatial structures of the pixels. In the following, we will introduce a new 2D tensor (a.k.a., order 2 tensor) representation, which we call the Global-Local representation. It is firstly proposed by (Chen et al., 2005), and later on advocated by (Hua et al., 2007).

Instead of using the raw images directly as the 2D tensor representation. The Global-Local representation firstly partitions the original raw face image into non-overlapping blocks. Following the raster scan order, each block is then raster-scanned as a column vector and concatenated together to form the new Global-Local representation. This transformation process is illustrated in Figure.1.

The biggest merit of the Global-Local representation is that it explores the spatial structure of the face image pixels in a good fashion. As we can clearly observe in Figure.1, the column vector of the Global-Local 2D tensor representation is the unfolded vectors from the local blocks of the original raw image. On the other hand, it is also easy to see that the row vector of the Global-Local representation is indeed the unfolded vector of smaller images down-sampled from the original image. Why it is better to perform discriminative subspace analysis on this Global-Local tensor representation?
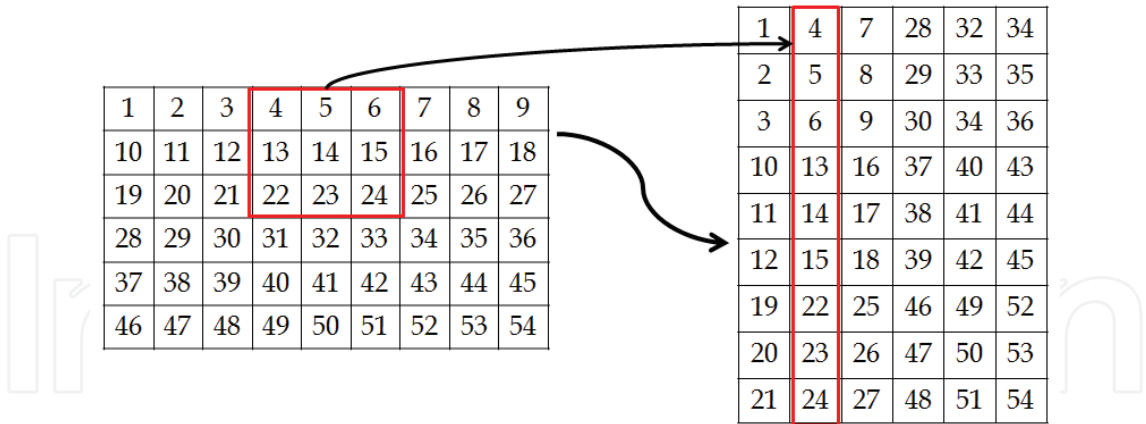
Fig. 1. Original 6×9 2D tensor (left side) and the Glocal-Local Tensor representation of 9×6 (right side) based on 3×3 local blocks.

Let us take a look of the operations of the left projections on the Global-Local tensor representation. By putting it back into the context of the original raw image, it is straightforward to see that the left projection is equivalent to convolute a local filter repeatedly on the different block partitions. Therefore, pursuing discriminative left projections is equivalent to identifying the most discriminative local filters for the original raw image.

On the other hand, the right projection is operating on the row vector of the Global-Local tensor representation. By putting it back into the context of the original raw image, the interpretation could be two-folds: First, by itself the projection is filtering on the down-sampled and shifted version of the original raw face image; on the other hand, coupling with the right projection, it selects which block partition we should weight the most to achieve the highest discriminative power.

Therefore, the combined interpretation of pursuing discriminative bi-linear projection with the Global-Local tensor representation is to seek for the most discriminative local filter and the best weighting scheme for the local pixel blocks. It is more sensible than using the raw face images directly as the 2D tensor representation. It is also clear that the Global-Local representation better utilized the spatial structure of the pixels on the face images.

In the rest of the Chapter, by default all the 2D tensors are with the Global-Local representation. We present here a discriminative orthogonal rank-one tensor decomposition method for face recognition, which is first proposed by (Hua et al., 2007).

4. Discriminative orthogonal rank-one tensor decomposition

In this section, we present the mathematic formulation of the discriminative orthogonal rank one tensor decomposition method followed by the detailed algorithm of how to pursue the tensor decomposition based on a set of labelled training data set. We present all the mathematic formulation under order n tensor but it should be just straightforward to derive from it for order 2 tensors.

We start from a set of training examples $\mathcal{T} = \{\mathbf{X}_i : \mathbf{X}_i \in \mathbf{R}^{n_1 \times n_2 \times \dots \times n_n}, i = 1,2,\dots,N\}$ with pair-wise labels $\mathcal{L} = \{l_{ij} : 1 \le i < j \le N, l_{ij} \in \{0,1\}\}$ where $l_{ij} = 1$ if $\mathbf{X}_i$ and $\mathbf{X}_j$ are in the same category (i.e., the faces of the same person under the context of face recognition), and $l_{ij} = 0$ if $\mathbf{X}_i$ and $\mathbf{X}_j$ are in different categories. We denote the k-nearest neighbour of the example $\mathbf{X}_i$

in the original input space to be $\mathcal{N}_k(\mathbf{X}_i)$. Then we define the positive label set and negative label set as $\mathcal{S} = \{(i,j): l_{ij} = 1, 1 \leq i < j \leq N, \mathbf{X}_i \in \mathcal{N}_k(\mathbf{X}_j) \text{ or } \mathbf{X}_j \in \mathcal{N}_k(\mathbf{X}_i)\}$ and $\mathcal{D} = \{(i,j): l_{ij} = 1, 1 \leq i < j \leq N, \mathbf{X}_i \in \mathcal{N}_k(\mathbf{X}_j) \text{ or } \mathbf{X}_j \in \mathcal{N}_k(\mathbf{X}_i)\}$, which are the k-nearest neighbour example pairs from the same or different categories, respectively.

For pursuing a discriminative embedding for face recognition, our objective here is to learn a set of orthogonal rank-one tensor projections $\{\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_K\}$ such that in the projective embedding space, the distance for those example pairs in $\mathcal{S}$ are minimized while the distance for those example pairs in $\mathcal{D}$ are maximized.

Following similar ideas as in (Duchene & Leclercq, 1988), we optimize a series of locally weighted discriminative cost function to build the discriminative embedding. More formally, suppose that we have already discriminatively pursued $k - 1$ orthogonal rank one tensor projections $\{\hat{P}^{(1)}, \hat{P}^{(2)}, \ldots, \hat{P}^{(k-1)}\}$, to pursue the kth rank one tensor projections, we solve for the following optimization problem,

$$\max_{\hat{P}^{(k)}} \frac{\sum_{(i,j)\in\mathcal{D}} \omega_{ij}(\hat{P}^{(k)} \circledast \mathbf{X}_i - \hat{P}^{(k)} \circledast \mathbf{X}_j)^2}{\sum_{(i,j)\in\mathcal{S}} \omega_{ij}(\hat{P}^{(k)} \circledast \mathbf{X}_i - \hat{P}^{(k)} \circledast \mathbf{X}_j)^2} \tag{9}$$

$$\text{s.t. } \hat{P}^{(k)} \perp \hat{P}^{(k-1)}, \hat{P}^{(k)} \perp \hat{P}^{(k-2)}, \ldots, \hat{P}^{(k)} \perp \hat{P}^{(1)} \tag{10}$$

where ω_{ij} is a weight assigned according to the importance of the example pair $\{\mathbf{X}_i, \mathbf{X}_j\}$. There are different strategies in setting the weight ω_{ij}. In our experiments, we adopted the most popular heat kernel weights, i.e., $\omega_{ij} = \exp\left\{-\left\|\mathbf{X}_i - \mathbf{X}_j\right\|_F^2/t\right\}$, where $\|\bullet\|_F$ denotes the Frobenius norm of matrices, and t is a constant heat factor. This weight setting induces heavy penalties to the cost function in Equation (9) for example pairs which are very close in the input space. One more thing to be noticed is that for $k = 1$, we only need to solve for the unconstrained optimization problem in Equation (9).

To solve for the constrained optimization problem in Equation (9~10), we are confronted by two difficulties: First, there is even no closed-form solution for the unconstrained optimization problem in Equation (9). Fortunately, it is well known that this unconstrained problem can be solved by using a sequential iterative optimization strategy. Second, it is in general difficult to keep both the rank-one and orthogonality properties. We address this issue by leveraging the sufficient and necessary conditions for orthogonal rank one tensors in Theorem 1.

In essence, Theorem 1 states that to make two rank one tensors to be orthogonal to each other, we only need to place the orthogonal constraints on one dimension of the rank-one tensors. Therefore, an equivalent set of constraints to the orthogonality constraints is

$$\exists \{j_l: l = 1, 2, \ldots, k-1; 1 \leq j_l \leq n\} \quad s.t. \quad \mathbf{p}_{j_{k-1}}^{(k)} \perp \mathbf{p}_{j_{k-1}}^{(k-1)}, \mathbf{p}_{j_{k-2}}^{(k)} \perp \mathbf{p}_{j_{k-2}}^{(k-2)}, \ldots, \mathbf{p}_{j_1}^{(k)} \perp \mathbf{p}_{j_1}^{(1)}, \tag{11}$$

where $\mathbf{p}_j^{(k)}$ indicates the projection vector corresponding to the jth dimension of the rank one tensor projection $\hat{P}^{(k)}$ which is of order n.

To ease the optimization process, we replace the constraints in Equation (11) with another set of stronger constraints, i.e.,

$$\exists \{j: 1 \leq j \leq n\} \quad s.t. \quad \mathbf{p}_j^{(k)} \perp \mathbf{p}_j^{(k-1)}, \mathbf{p}_j^{(k)} \perp \mathbf{p}_j^{(k-2)}, \ldots, \mathbf{p}_j^{(k)} \perp \mathbf{p}_j^{(1)} \tag{12}$$

These constraints are stronger in the sense that it requires all the different j_l in Equation (11) to be same value. It is just trying to put all orthogonal constraints on one dimension of the rank-one tensor projections. With the sufficient condition to ensure the orthogonal property for the rank-one projections in Equation (12), we proceed to derive the solution for the constrained optimization problem in Equation (9~10).

As we have mentioned beforehand, the unconstrained optimization problem in Equation (9) is usually solved numerically in a sequential iterative fashion. That is, at each iteration, we fix $\hat{P}_{-i}^{(k)} = \{\mathbf{p}_1^{(k)}, \mathbf{p}_2^{(k)}, \dots, \mathbf{p}_{i-1}^{(k)}, \mathbf{p}_{i+1}^{(k)}, \dots, \mathbf{p}_n^{(k)}\}$ for one of the $1 \leq i \leq n$, and optimize Equation (9) with respect to $\mathbf{p}_i^{(k)}$. As a matter of fact, once we fixed $\hat{P}_{-i}^{(k)}$, the optimization problem boils down to a problem in a vector space of dimension n_i. To simplify the notation, we denote

$$\mathbf{y}^{(k,i)} = \mathbf{X} \times_1 \mathbf{p}_1^{(k)} \times_2 \mathbf{p}_2^{(k)} \times_3 \dots \times_{i-1} \mathbf{p}_{i-1}^{(k)} \times_{i+1} \mathbf{p}_{i+1}^{(k)} \times_{i+2} \dots \times_n \mathbf{p}_n^{(k)} \overset{\text{def}}{=} \mathbf{X} \circledast \hat{P}_{-i}^{(k)} \tag{13}$$

which is an n_i dimensional vector. Then it is easy to figure out that the optimization problem in Equation (9) boils down to the following problem

$$\arg\max_{\mathbf{p}_i^{(k)}} \frac{\mathbf{p}_i^{(k)^{\mathrm{T}}} \mathbf{A}_d^i \mathbf{p}_i^{(k)}}{\mathbf{p}_i^{(k)^{\mathrm{T}}} \mathbf{A}_s^i \mathbf{p}_i^{(k)}} \tag{14}$$

where

$$\mathbf{A}_d^i = \sum_{(s,t)\in\mathcal{D}} \omega_{st} \left(\mathbf{y}_s^{(k,i)} - \mathbf{y}_t^{(k,i)}\right) \left(\mathbf{y}_s^{(k,i)} - \mathbf{y}_t^{(k,i)}\right)^T \tag{15}$$

$$\mathbf{A}_s^i = \sum_{(s,t)\in\mathcal{S}} \omega_{st} \left(\mathbf{y}_s^{(k,i)} - \mathbf{y}_t^{(k,i)}\right) \left(\mathbf{y}_s^{(k,i)} - \mathbf{y}_t^{(k,i)}\right)^T \tag{16}$$

$$\mathbf{y}_o^{(k,i)} = \mathbf{X}_o \circledast \hat{P}_{-i}^{(k)}. \tag{17}$$

It is also well known that the solution to the unconstrained optimization problem in Equation (14) could be obtained by solving a generalized eigenvalue system, i.e.,

$$\mathbf{A}_d^i \mathbf{p} = \lambda \mathbf{A}_s^i \mathbf{p} \tag{18}$$

and the optimal $\mathbf{p}_i^{(k)*}$ is the eigenvector associated with the largest eigenvalue. Equation (15) is solved iteratively over $i = 1, 2, \dots, n$ until convergence. The converged output $\hat{P}^{(k)*} = \{\mathbf{p}_1^{(k)*}, \mathbf{p}_2^{(k)*}, \dots, \mathbf{p}_i^{(k)*}, \dots, \mathbf{p}_n^{(k)*}\}$ is regarded to the optimal solution to the unconstrained optimization problem of Equation (9). It only guarantees a local optimal solution, though.

But we are missing the orthogonal constraints Equation (10) here. As we have discussed, the constraints in Equation (12) is a sufficient condition for the constraint in Equation (10). So we need to ensure the constraints in Equation (12). It immediately implies that we only need to ensure the orthogonality for one of the dimension j during the sequential iterative optimization process to ensure the orthogonality of the tensor projections.

That is to say, for $i \neq j$, we only need to solve for an unconstrained optimization problem in Equation (14). But for $i = j$, we essentially need to solve for the following constrained optimization problem,

$$\arg\max_{\mathbf{p}_i^{(k)}} \frac{\mathbf{p}_j^{(k)^{\mathrm{T}}} \mathbf{A}_d^j \mathbf{p}_j^{(k)}}{\mathbf{p}_j^{(k)^{\mathrm{T}}} \mathbf{A}_s^j \mathbf{p}_j^{(k)}} \tag{19}$$

$$s.t. \ \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(k-1)} = 0, \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(k-2)} = 0, \dots, \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(1)} = 0 \tag{20}$$

It is easy to see that it is equivalent to solve for the following constrained optimization problem, i.e.,

$$\arg\max_{\mathbf{p}_i^{(k)}} \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} \tag{21}$$

$$s.t. \quad \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_s^j \mathbf{p}_j^{(k)} = 1, \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(k-1)} = 0, \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(k-2)} = 0, \dots, \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(1)} = 0. \tag{22}$$

For the constrained optimization problem in Equation (21~22), we show here that the optimal solution can be obtained by solving for the following eigenvalue problem:

$$\widetilde{\mathcal{M}} \mathbf{p}_j^{(k)} = \left(\mathbf{I} - \left(\mathbf{A}_s^j\right)^{-1} \mathcal{A} \mathcal{B}^{-1} \mathcal{A}^T\right) \mathbf{A}_s^{j-1} \mathbf{A}_d^j \mathbf{p}_j^{(k)} = \lambda \mathbf{p}_j^{(k)} \tag{23}$$

where

$$\mathcal{A} = \left[\mathbf{p}_j^{(1)}, \mathbf{p}_j^{(1)}, \dots, \mathbf{p}_j^{(k-1)}\right] \tag{24}$$

$$\mathcal{B} = \mathcal{A}^T \mathbf{A}_s^{j-1} \mathcal{A}. \tag{25}$$

The optimal $\mathbf{p}_j^{(k)*}$ is the eigenvector corresponding to the largest eigenvalue of $\widetilde{\mathcal{M}}$. Following similar steps as shown in (Hua et al., 2007; Duchene & Leclercq, 1988), in the following we demonstrate how we derive the solution presented in Equation (23).

We firstly formulate the Lagrangian multipliers out of the constrained optimization problem in Equation (21~23), i.e.,

$$L\left(\mathbf{p}_j^{(k)}, \lambda, \mu_1, \mu_2, \dots, \mu_{k-1}\right) = \mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} - \lambda \left(\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_s^j \mathbf{p}_j^{(k)} - 1\right)$$

$$-\mu_{k-1}\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(k-1)} - \dots - \mu_2\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(2)} - \mu_1\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{p}_j^{(1)}. \tag{26}$$

Take the derivative of $L\left(\mathbf{p}_j^{(k)}, \lambda, \mu_1, \mu_2, \dots, \mu_{k-1}\right)$ with respect to $\mathbf{p}_j^{(k)}$, and set it to zero, we have

$$\frac{\partial L\left(\mathbf{p}_j^{(k)}, \lambda, \mu_1, \mu_2, \dots, \mu_{k-1}\right)}{\partial \mathbf{p}_j^{(k)}} = 2\mathbf{A}_d^j \mathbf{p}_j^{(k)} - 2\lambda \mathbf{A}_s^j \mathbf{p}_j^{(k)} - \mu_{k-1}\mathbf{p}_j^{(k-1)} - \dots - \mu_1\mathbf{p}_j^{(1)} = 0 \tag{27}$$

Left multiply both side of Equation (27) by $\mathbf{p}_j^{(k)^\mathrm{T}}$, we immediately have

$$\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} - \lambda \left(\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_s^j \mathbf{p}_j^{(k)} - 1\right) = 0. \tag{28}$$

We have

$$\lambda = \frac{\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_d^j \mathbf{p}_j^{(k)}}{\mathbf{p}_j^{(k)^\mathrm{T}} \mathbf{A}_s^j \mathbf{p}_j^{(k)}} \tag{29}$$

which is exactly the quantity we want to maximize in Equation (19). Multiply both side of Equation (29) by $\mathbf{p}_j^{(l)^{\mathrm{T}}} \mathbf{A}_s^{j^{-1}}$ for $l = 1,2, \dots, k-1$, and with easy manipulation, we obtain a set of $k-1$ equations, i.e.,

$$\sum_{m=1}^{k-1} \mu_m \mathbf{p}_j^{(1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{p}_j^{(m)} = 2\mathbf{p}_j^{(1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} \tag{30}$$

$$\sum_{m=1}^{k-1} \mu_m \mathbf{p}_j^{(2)^T} \mathbf{A}_s^{j^{-1}} \mathbf{p}_j^{(m)} = 2\mathbf{p}_j^{(2)^T} \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} \tag{31}$$

$$\dots \dots \dots$$

$$\sum_{m=1}^{k-1} \mu_m \mathbf{p}_j^{(k-1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{p}_j^{(m)} = 2\mathbf{p}_j^{(k-1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)}. \tag{32}$$

We can write Equation (30~32) more concisely in matrix form as

$$\begin{bmatrix} \mathbf{p}_j^{(1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{p}_j^{(1)} & \cdots & \mathbf{p}_j^{(1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{p}_j^{(k-1)} \\ \vdots & \ddots & \vdots \\ \mathbf{p}_j^{(k-1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{p}_j^{(1)} & \cdots & \mathbf{p}_j^{(k-1)^T} \mathbf{A}_s^{j^{-1}} \mathbf{p}_j^{(k-1)} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{k-1} \end{bmatrix} = 2 \begin{bmatrix} \mathbf{p}_j^{(1)^T} \\ \mathbf{p}_j^{(2)^T} \\ \vdots \\ \mathbf{p}_j^{(k-1)^T} \end{bmatrix} \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)}. \tag{33}$$

We can further simplify Equation (33) to be

$$\begin{bmatrix} \mathbf{p}_j^{(1)^T} \\ \mathbf{p}_j^{(2)^T} \\ \vdots \\ \mathbf{p}_j^{(k-1)^T} \end{bmatrix} \mathbf{A}_s^{j^{-1}} \begin{bmatrix} \mathbf{p}_j^{(1)} & \mathbf{p}_j^{(2)} & \cdots & \mathbf{p}_j^{(k-1)} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{k-1} \end{bmatrix} = 2 \begin{bmatrix} \mathbf{p}_j^{(1)^T} \\ \mathbf{p}_j^{(2)^T} \\ \vdots \\ \mathbf{p}_j^{(k-1)^T} \end{bmatrix} \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} \tag{34}$$

Denote $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_{k-1}]^T$, and use the notation in Equation (24~25), we can rewrite Equation (34) to be

$$\mathcal{B}\boldsymbol{\mu} = \mathcal{A}^T \left(\mathbf{A}_s^j\right)^{-1} \mathcal{A}\boldsymbol{\mu} = 2\mathcal{A}^T \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} \tag{35}$$

Therefore, we have

$$\boldsymbol{\mu} = 2\mathcal{B}^{-1} \mathcal{A}^T \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} \tag{36}$$

Multiply both side of Equation (27) by $\mathbf{A}_s^{j^{-1}}$ and rearrange it to be in matrix form, we can easily obtain

$$2\mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} - 2\lambda \mathbf{p}_j^{(k)} - \mathbf{A}_s^{j^{-1}} \mathcal{A}\boldsymbol{\mu} = 0. \tag{37}$$

Embedding Equation (36) into Equation (37), we obtain

$$2\mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} - 2\lambda \mathbf{p}_j^{(k)} - 2\mathbf{A}_s^{j^{-1}} \mathcal{A}\mathcal{B}^{-1} \mathcal{A}^T \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} = 0. \tag{38}$$

$Input$: $\mathcal{T} = \{\mathbf{X}_i : \mathbf{X}_i \in \mathbf{R}^{n_1 \times n_2 \times \dots \times n_n}, i = 1, 2, \dots, N\}$

$\mathcal{S} = \{(i, j) : l_{ij} = 1, 1 \le i < j \le N, \mathbf{X}_i \in \mathcal{N}_k(\mathbf{X}_j) \text{ or } \mathbf{X}_j \in \mathcal{N}_k(\mathbf{X}_i)\}$

$\mathcal{D} = \{(i, j) : l_{ij} = 1, 1 \le i < j \le N, \mathbf{X}_i \in \mathcal{N}_k(\mathbf{X}_j) \text{ or } \mathbf{X}_j \in \mathcal{N}_k(\mathbf{X}_i)\}$

$Output$: $\hat{\mathcal{P}} = \{\hat{P}^{(1)}, \hat{P}^{(2)}, \dots, \hat{P}^{(K)}\}$, a set of K discrimiantive rank one tensor projections.

1. Initialize $k = 0$, and randomly initialize each vector $\mathbf{p}_i^{(k)}$ as a normal vector for $i = 1, 2, \dots, n$. Then sequentially and iteratively solve for the unconstrained discriminative eigenvalue problem in Equation (18) until convergence to obtain the first discriminative rank-one tensor projection $\hat{P}^{(1)}$. Set $k = k + 1$.

2. Randomly initialize each vector $\mathbf{p}_i^{(k)}$ as a normal vector for $i = 1, 2, \dots, n$. Then randomly generate a number j, such that $0 \le j \le n \,\&\, c_k(j) < n_j$ where $c_k(j)$ indicates the number of times that dimension j was picked up prior to this step k.

   2a. For each $i = j, 1, 2, \dots, j - 1, j + 1, \dots, n$, fix all the other projection vectors except $\mathbf{p}_i^{(k)}$, i.e., $\hat{P}_{-i}^{(k)} = \{\mathbf{p}_1^{(k)}, \mathbf{p}_2^{(k)}, \dots, \mathbf{p}_{i-1}^{(k)}, \mathbf{p}_{i+1}^{(k)}, \dots, \mathbf{p}_n^{(k)}\}$. If $i = j$, then solve for the eigenvalue system in Equation (23) to update $\mathbf{p}_i^{(k)}$. Otherwise, solve for the eigenvalue system in Equation (18) to update $\mathbf{p}_i^{(k)}$. Normalize $\mathbf{p}_i^{(k)}$ after the update.

   2b. Repeat step 2a until convergence, we obtained the kth discriminative rank-one projection $\hat{P}^{(k)}$. Go to step 3.

3. Set $k = k + 1$, if $k < K$, repeat step 2. Otherwise output the final set of discriminative rank-one tensor projection: $\hat{\mathcal{P}} = \{\hat{P}^{(1)}, \hat{P}^{(2)}, \dots, \hat{P}^{(K)}\}$.

Fig. 2. Orthogonal rank-one tensor discriminative decomposition.

From Equation (29), it is straightforward to have

$$\left(\mathbf{I} - \mathbf{A}_s^{j^{-1}} \mathcal{A} \mathcal{B}^{-1} \mathcal{A}^T\right) \mathbf{A}_s^{j^{-1}} \mathbf{A}_d^j \mathbf{p}_j^{(k)} = \widetilde{\mathcal{M}} \mathbf{p}_j^{(k)} = \lambda \mathbf{p}_j^{(k)}. \tag{39}$$

Since λ is exactly the quantity we want to maximize, we have the conclusion that the optimal $\mathbf{p}_j^{(k)^*}$ for the constrained optimization problem in Equation (19~20) or Equation (21~22) is the eigenvector corresponding to the largest eigenvalue of the matrix $\widetilde{\mathcal{M}}$.

With all the analysis above, we summarize here a sequential iterative optimization scheme for solving the constrained optimization problem in Equation (9~10), namely discriminative orthogonal rank-one tensor decomposition, as shown in Figure 2. Such a discriminative orthogonal rank-one tensor decomposition method is firstly presented in (Hua et al., 2007). Note when choosing the dimension to reinforce the orthogonal constraint in Step 2 of Figure 2, we cannot choose the same dimension j for more than n_j times because there are at most n_j vector can be orthogonal to each other in a n_j dimensional vector space. In the next section, we present some experimental results on face recognition using our method, and compare them with the state-of-the-art discriminative embedding methods for face recognition, with either vector or tensor based representation.

Method\Dataset	Recognition Rate (%)/Dimension of the embedding space			
	Yale	ORL	YaleB	PIE
SSD baseline	54.4%/1024	88.1%/1024	65.4%/1024	62.1%/1024
PCA	54.8%/ 71	88.1%/ 189	65.4%/ 780	62.1%/1023
LDA	77.5%/ 14	93.9%/ 39	81.3%/ 37	89.1%/ 67
LPP	78.3%/ 14	93.7%/ 39	86.4%/ 76	**89.2%/ 86**
Tensor LPP	76.4%/ 35	**95.8%/ 71**	**92.4%/ 311**	**90.3%/ 68**
OLPP	**82.1%/ 14**	**96.6%/ 41**	**94.3%/ 241**	**93.6%/381**
RPAM	**79.1%/242**	92.0%/ 219	**92.4%/ 389**	**89.8%/399**
2DLDE$_{4\times2}$	**80.7%/ 113**	**95.5%/ 87**	**90.2%/ 88**	88.0%/ 104
ORO	70.2%/ 32	92.8%/ 30	88.1%/ 32	88.1%/ 31
ORO$_{4\times4}$	**80.8%/ 53**	**95.2%/ 58**	89.1%/ 53	**91.5%/ 49**
ORO$_{4\times2}$	**86.8%/ 94**	**97.0%/ 105**	**91.0%/ 108**	**93.6%/ 73**
	(**82.4%/ 14**)	(**95.0%/ 41**)	-------	-------

Table 1. Face recognition results on Yale, ORL, YaleB and PIE.

5. Experiments and discussions

The proposed method of using discriminative rank-one tensor projections with Global-Local tensor representation are extensively tested on four widely used benchmark face recognition datasets including the Yale dataset (Belhumeur et al., 1997), the Olivetti Research Laboratory (ORL) database (Samaria & Harter 1994), the extended Yale face database B dataset (Georghiades et al., 2001), and the CMU PIE dataset (Sim et al., 2003). We call them Yale, ORL, YaleB, and PIE, respectively.

In all the dataset, we crop the grey scale face images, and align all face images based on their eye positions. The aligned face image is then resized to be 32×32 images. No other pre-processing on the image is performed. For each dataset, we randomly split the dataset into training and testing dataset. The average performance over several random splits is reported. Except for Yale dataset, on which we report results with 20 random splits, the results from 50 random splits are aggregated for all the other three dataset. All the results we discuss here are summarized from (Hua et al., 2007). In our experiments, the face recognition is performed based on a 1-Nearest Neighbour classifier based on the Euclidean distance on the embedding space.
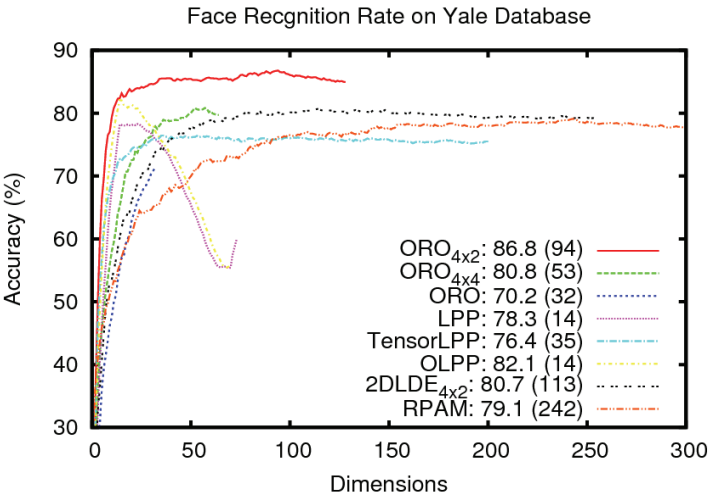

Fig. 3. Face recognition results on the Yale data set (recognition rate v.s. dimensionality).

The discriminative orthogonal rank-one tensor decomposition method is tested with 3 different settings by performing it on: a.) raw image 2D tensors representations; b.) Global-Local tensor representation based on 4×2 block partitions; and c.) Global-Local tensor representation based on 4×4 block partitions. We name them ORO, $ORO_{4\times2}$, and $ORO_{4\times4}$, respectively. We have compared the results from these three settings with almost all the state-of-the-art linear embedding methods such as PCA (Turk & Pentlend, 1991), LDA (Belhumeur et al., 1997), LPP (He et al., 2005a), tensor LPP (He et al, 2005b), orthogonal LPP (Cai et al., 2006), the two dimensional local discriminative embedding with Global-Local representation based on 4×2 blocks ($2DLDE_{4\times2}$) from (Chen et al., 2005), and the Rank-one projection with adaptive margins (RPAM) (Xu et al., 2006).

The recognition accuracies of all the different methods are presented in Table 1. As a baseline, we also present the results of using SSD in the raw image space. For each dataset, the top 5 performed methods are highlighted in the table. In the following subsections, we will discuss in more details of the results dataset by dataset.

Yale Data Set: The Yale data set is indeed a very small face benchmark. It contains 165 faces of 15 different individuals with different facial expressions. The results of the different methods running on this data set are presented in the first column of Table 1. The results reported are the average accuracy over 20 random splits of the data set, with 5 from each person for training and the rest for testing. Therefore, each split utilizes 55 faces for training and 110 for testing.

As we can clearly observe, $ORO_{4\times2}$ achieves the best recognition accuracy of 86.8% with 94 dimensions. Its performance is significantly better than all the other methods. In Figure 3, we present the recognition rates of different methods versus the number of dimensionality of the embedding space. It clearly shows that $ORO_{4\times2}$ outperforms all the other methods. Interestingly, when it goes beyond dimension 14, which is the maximum number of projections LDA can pursue (because there are only 15 different subjects), the recognition accuracy for $ORO_{4\times2}$ continues to go up. The recognition accuracies for both the LPP and the OLPP drop rapidly.

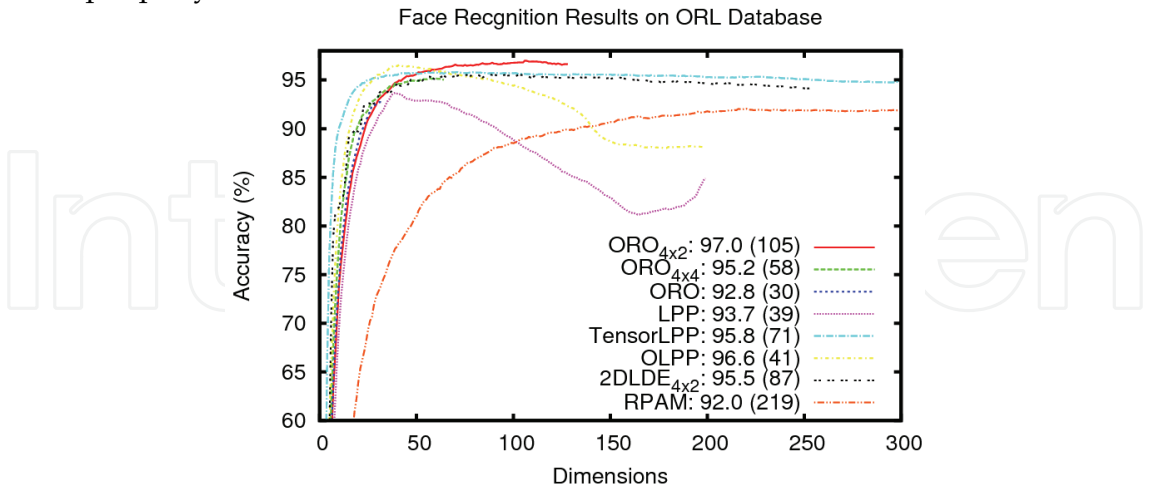Face Recgnition Results on ORL Database

Fig. 4. Face recognition results on the ORL data set (recognition rate v.s. dimensionality).

It is also observed that ORO did not perform as well as the Tensor LPP and RPAM methods. Our intuition is that for small training example set, the orthogonal regularization on the 32×32 rank one tensor projections is too strong. Moreover each rank-one projection only has 64 parameters, which significantly limits the capacity of the rank-one projection.

Without the orthogonal constraints, the Tensor LPP method and RPAM method are able to leverage some additional capacities to achieve higher recognition rate. Last but not least, the effective-ness of the orthogonal constraint regularization can be understood by comparing the result of $ORO_{4\times2}$ with that of $2DLDE_{4\times2}$ since the only difference of the two methods are the orthogonal regularization.

ORL Dataset: the ORL dataset has 40 different subjects. Each has 10 different faces which amount to a total of 400 faces. For each subject, the 10 different faces are taken at different time, under different lighting conditions, and with different facial expressions. In our experiments, 5 images are selected for each person to form the 200 training images, and the rest 200 images are used for testing purpose. The reported results are the aggregated results over 50 random splits.

Again, $ORO_{4\times2}$ leads all the other method, which achieves a recognition rate of 97% with 105 dimensions. This is followed by OLPP with a recognition rate of 96.6% with 41 dimensions. $ORO_{4\times2}$ with 41 dimensions achieves an accuracy of 95%, which is inferior to OLPP. But it is still better than PCA, LDA, LPP and RPAM. It is interesting to observe that with increased number of training examples compared with the Yale data set, the recognition rate of RPAM with 218 dimensions cannot beat that of ORO with only 32 dimensions. Assuming the adaptive margin step poses positive effects, it indicates that with the increased number of training examples, the orthogonal constraint really improves the ability for generalization for the learned rank-one tensor projections. We plot the recognition rate versus dimensionality of the embedding space for all the different methods in Figure 4.

YaleB Dataset: The YaleB dataset contains 21888 face images of 38 different persons under 9 poses and 64 illumination conditions. We choose the subset of 2432 nearly frontal faces (i.e., 64 face images per person). In our evaluation we randomly choose 20 images per person for
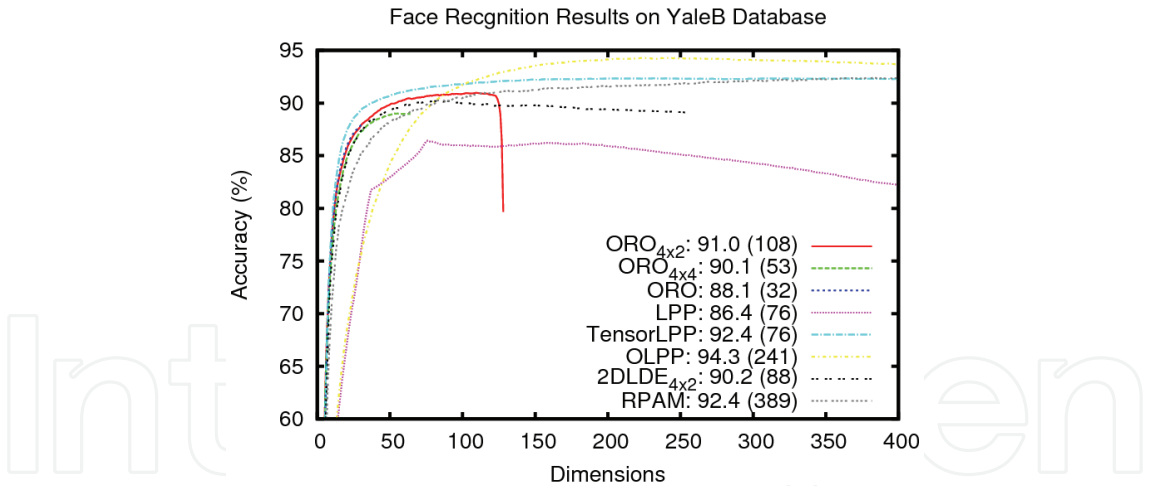

Fig. 5. Face recognition results on the YaleB data set (recognition rate v.s. dimensionality).

training and the rest for testing. Hence there are 760 training face images and 1672 testing faces. This training set is of medium size compared with the raw image dimensionality 1024. We report the results averaged over 50 random splits in the third column of Table 1. The recognition accuracy of $ORO_{4\times2}$ is 91.0% with 108 dimensions, which is better than LDA, lPP and $2DLDE_{4\times2}$, and inferior yet comparable to RPAM, Tensor LPP and OLPP. RPAM is obviously benefiting from the adaptive margin step. Moreover, with more training data, the negative effect of high dimensionality is less severe and thus OLPP may achieve better results. Again, we plot the recognition rate versus dimensionality of all the different methods on this dataset in Figure 5.

PIE Dataset: The PIE database contains 41368 face images of 68 people, which are taken under 13 poses, 43 illumination conditions, and 4 expressions). We use the images of 5 nearly front poses (C05, C07, C09, C27, C29) under all illumination conditions and expressions. This forms a subset of 11560 face images with 170 images per person. For each run, 30 images are randomly picked up for training and the rest 120 images per person are used for testing. Again, the average recognition rate over 50 different runs is summarized in the fourth column of Table 1.

Both the $ORO_{4\times2}$ and OLPP achieves the highest recognition rate of 93.6%. But $ORO_{4\times2}$ achieves this performance using only 73 dimensions while OLPP needs to pick up as high as 381 projection vectors. The red curve in Figure 6 shows how $ORO_{4\times2}$ can greedily pursue the smallest but most discriminative set of orthogonal rank-one projections to achieve the highest recognition rate.

6. Remarks

We highlight some of our general remarks about the performance of the discriminative rank-one tensor projections on the task of face recognition.

- First of all, it is noted in our experiments that the discriminative power (i.e., the largest eigenvalue corresponding to the linear system defined in either Equation (18) or Equation (23)) of consecutively pursued orthogonal rank-one projections is not monotonically decreasing. Therefore, after the final solution set was obtained, we need to sort these orthogonal rank-one tensor projections by their discriminative powers and pick up the top K ones to form the discriminative embedding for the face recognition task.
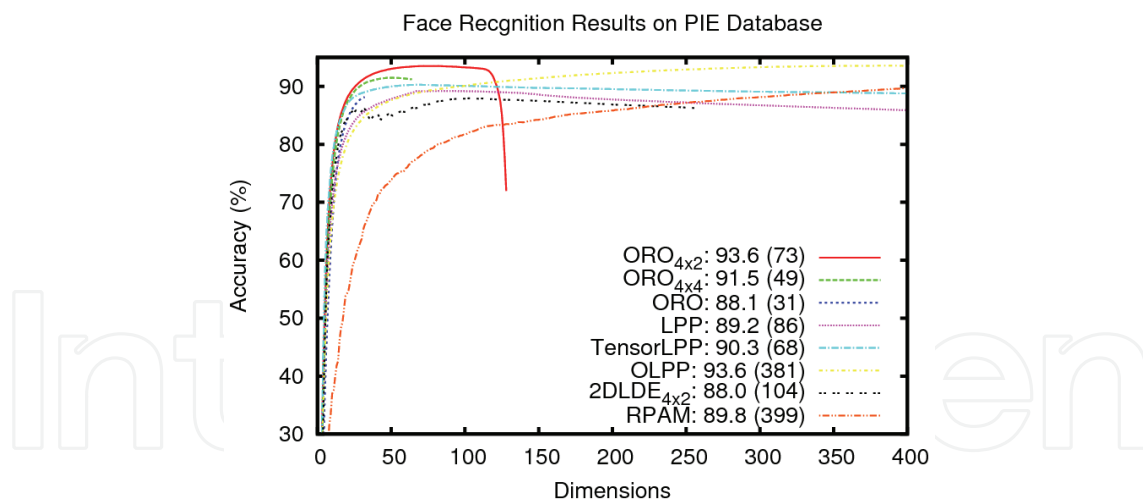

Fig. 6. Face recognition results on the PIE data set (recognition rate v.s. dimensionality).

- As shown in Figure (5~6), on the YaleB and PIE datasets, adding in the last several orthogonal rank-one projections obtained by $ORO_{4\times2}$ dramatically degrades the recognition accuracy. In this case the orthogonal regularization forces these last projections to preserve only non-discriminative information.
- The performance of ORO is limited by the number of orthogonal rank one projections we can obtain from the algorithm presented in Figure 2. However, on YaleB, it achieves the error rate of 11.9% with only 32 dimensions, which is much better than LDA (18.7%

with 37 dimensions) and LPP (13.6% with 76 dimensions). This may be partially due to the tensor based representation, which suffers less from the curse-of-dimensionality.

- The Global-Local tensor representation in general gives an significant boost to the performance. For example, the two methods $ORO_{4\times2}$ and $ORO_{4\times4}$, which adopted the Global-Local tensor representation, are consistently performing better across all the four datasets than ORO, which adopted the naive tensor representation of raw images.
- Posing orthogonal constraints on the discriminative rank-one tensor projections in general helps to improve the performance. This conclusion comes from comparing the recognition results between $ORO_{4\times2}$ and $2DLDE_{4\times2}$. $ORO_{4\times2}$ consistently achieves better recognition accuracy than $2DLDE_{4\times2}$ across all the four face benchmark.
- Overall, the two orthogonal constrained algorithms, $ORO_{4\times2}$ and OLPP achieve the best recognition rate. $ORO_{4\times2}$ outperforms OLPP on Yale and ORL, and achieves equivalent performance to that of OLPP on PIE. It is only inferior to OLPP on the YaleB dataset.
- RPAM (Xu et al., 2006) tends to require more projections to achieve a good performance. This may be due to the adaptive margin step, which seems to be effective according to our experiments.
- On small or medium size face datasets such as Yale and ORL, the discriminative orthogonal rank-one tensor projection method outperforms the other state-of-the-art discriminative embedding methods. On larger size database such as YaleB or PIE, it achieves comparable results to the best state-of-the-art, but uses much less number of projections. This is a very interesting phenomenon we observe. It surely makes it more scalable for face recognition on larger scale face databases.

Nevertheless, further investigation and consolidation of the remarks we summarized above is definitely beneficial to have a deeper understanding of the behaviour of the discriminative rank-one tensor decomposition method presented in this Chapter.

7. Conclusions

This Chapter illustrated two types of regularization methods recently developed in the computer vision literature for robust face recognition (Hua et al, 2007). The first regularization method is a new tensor representation of face images, which we call Global-Local tensor representation. It enables the successive discriminative embedding analysis to better leverage the geometric structure of the face image pixels. It also reinforces physically meaningful interpretation of the different dimensions of the tensor projections.

The second type of regularization method is an orthogonal constraint on discriminative rank-one tensor projections. We reveal a nice property of orthogonal rank-one tensors, which enables a fairly simple scheme to reinforce the orthogonality of the different rank-one projections. A novel, simple yet effective sequential iterative optimization algorithm is proposed to pursue a set of orthogonal rank-one tensor projections for face recognition.

By combining the two regularization methods, our extensive experiments demonstrate that it outperforms previous discriminative embedding methods for face recognition on small scale face databases. When dealing with larger face databases, it achieves comparable results to the best state-of-the-art, but results in more compact embeddings. In other words, it achieves comparable results to the best in the literature while uses much less number projections. This makes it far more efficient to handle larger face databases, in terms of both memory usage and recognition speed.

8. References

Duchene, J. & Leclercq S. (1988). An optimal transformation for discriminant and principal component analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.10, No.6, November 1988 (978–983).

Turk M. A. & Pentland A. P. (1991). Face recognition using eigenfaces. *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 586–591, June 1991.

Samaria, F. & Harter A. (1994). Parameterization of a stochastic model for human face identification. *Proceedings of IEEE Workshop on Applications of Computer Vision*, pp.138–142, Sarasota, FL, USA, December 1994.

Belhumeur, P. N.; Hespanha J. P. & Kriegman D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 19, No.7, July 1997 (711–720). Special Issue on Face Recognition.

Kolda T. G.(2001). Orthogonal tensor decompositions. *SIAM Journal on Matrix Analysis and Applications*, Vol.23, NO.1, January, 2001 (243–257).

Georghiades, A. S.; Belhumeur, P. N., & Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelli.*, Vol.23, No.6, June 2001 (643–660).

Sim, T. ; Baker, S. ; & Bsat, M. (2003). The cmu pose, illumination, and expression database. *IEEE Trans. on Pattern Anal. Mach. Intelli.*, Vol.25, No.12, December 2003 (1615–1618).

Chen, H.-T.; Liu, T.-L. & Fuh, C.-S. (2005). Learning effective image metrics from few pairwise examples. *Proceedings of IEEE International Conf. on Computer Vision*, pp. 1371–1378, Beijing, China, Octobor 2005.

He, X.F.; Yan S.C.; Hu, X.; Niyogi, P. & Zhang, H.J. (2005a). Face recognition using laplacianfaces. *IEEE Transaction on Pattern Anal Mach. Intell.*, Vol.27, NO.3, March, 2005 (328–340).

He X.F.; Cai D.; & Niyogi P. (2005b). Tensor subspace analysis. *Proceedings of Advances in Neural Information Processing Systems*, Vol18, Vancouver, Canada, December 2005.

Xu D. ; Lin S. ; Yan S.C. & Tang X. (2006). Rank-one projections with adaptive margins for face recognition. *Proceedings of IEEE Conf. on Computer Vision and Patter Recognition*, Vol.1, pp. 175–181, New York City, NY, June 2006.

Cai, D.; He, X.F.; Han, J.; & Zhang, H.-J. (2006). Orthogonal laplacianfaces for face recognition. *IEEE Trans. on Image Processing*, Vol. 15, No.11, November 2006 (3608–3614).

Hua, G.; Viola, P. & Drucker, S. (2007). Face Recognition using Discriminatively Trained Orthogonal Rank One Tensor Projections, *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, Minneaplois*, MN, 2007.

Yan, S.C. ; Xu, D. ; Zhang, B. ; Zhang, H.J. ; Yang, Q. & Lin, S. (2007). Graph Embedding and Extensions: A General Framework for Dimensionality Reduction, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.29, No.1, January, 2007 (40-51)

Recent Advances in Face Recognition

Edited by Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett

ISBN 978-953-7619-34-3

Hard cover, 236 pages

Publisher InTech

Published online 01, June, 2008

Published in print edition June, 2008

The main idea and the driver of further research in the area of face recognition are security applications and human-computer interaction. Face recognition represents an intuitive and non-intrusive method of recognizing people and this is why it became one of three identification methods used in e-passports and a biometric of choice for many other security applications. This goal of this book is to provide the reader with the most up to date research performed in automatic face recognition. The chapters presented use innovative approaches to deal with a wide variety of unsolved issues.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Gang Hua (2008). Face Recognition by Discriminative Orthogonal Rank-one Tensor Decomposition, Recent Advances in Face Recognition, Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett (Ed.), ISBN: 978-953-7619-34-3, InTech, Available from:
http://www.intechopen.com/books/recent_advances_in_face_recognition/face_recognition_by_discriminative_orthogonal_rank-one_tensor_decomposition

INTECH
open science | open minds