Research Article

Incremental Discriminant Analysis in Tensor Space

Liu Chang,^{1,2} Zhao Weidong,^{1,2} Yan Tao,^{1,2} Pu Qiang,^{1,2} and Du Xiaodan^{1,2}

¹College of Information Science and Technology, Chengdu University, Chengdu 610106, China

²Key Laboratory of Pattern Recognition and Intelligent Information Processing in Sichuan, Chengdu 610106, China

Correspondence should be addressed to Du Xiaodan; duxiaodandan@cdu.edu.cn

Received 20 September 2014; Revised 4 January 2015; Accepted 8 February 2015

Academic Editor: Rahib H. Abiyev

Copyright © 2015 Liu Chang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To study incremental machine learning in tensor space, this paper proposes incremental tensor discriminant analysis. The algorithm employs tensor representation to carry on discriminant analysis and combine incremental learning to alleviate the computational cost. This paper proves that the algorithm can be unified into the graph framework theoretically and analyzes the time and space complexity in detail. The experiments on facial image detection have shown that the algorithm not only achieves sound performance compared with other algorithms, but also reduces the computational issues apparently.

1. Introduction

Nowadays, increasing amounts of data in the field of industrial, economic, medical, and other application areas, such as signals, measurements, images, and videos, are becoming available due to the development of computer technology. In order to excavate the hidden information in the data implicitly describing underlying processes or structures, advanced intelligent tools are proposed. However, since the stochastic nature of the processes and their measurement, structure in this data is mostly collected with noise. Consequently, it is reasonable to seek robust and adaptive tools that can cope with this nature.

Computational intelligence techniques have been investigated to answer this need. These techniques have been concerned with reproducing the abilities of human brains. Machine learning techniques exactly imitate the learning procedure of human, which construct learning model based on example data and use that to make predictions and decisions. However, due to the noise in data, it is important to construct efficient learning model to help sift useful information from the noise.

In regards to this, machine learning algorithms project high-dimensional data into low-dimensional feature space to make their low-features as separable as possible. Generally, they are classified into two categories: supervised learning and unsupervised learning. The essential difference

between supervised learning and unsupervised learning is that whether the class information is considered. Generally speaking, the recognition performance of supervised learning is superior to that of unsupervised learning. As a classical machine learning algorithm, linear discriminate analysis (LDA) [1, 2] seeks optimal discriminative vectors to maximize the interclass scatter matrix and to minimize the intraclass scatter matrix. A large number of research works have shown the predominant advantage of LDA in various applications.

It is worth noting that traditional LDA is based on vector model. It requires all data being vectorized before learning. Actually, high-dimensional image data is structured data; the vectorization operation will break the correlation relationship of different pixels. Furthermore, the vectorization operation also is easy to result in the curse of dimensionality problem. As a result, machine learning algorithms [3–11] based on tensor algebra are investigated. These algorithms consider high-dimensional image as a high order tensor and introduce tensor algebra to analyze tensor data. Tensor representation not only is helpful to preserve the structure of high-dimensional image, but also serves as an effective way to avoid the curse of dimensionality problem. To unify all machine learning algorithms, [12] proposes the graph embedding framework. Under this graph embedding framework, two kinds of projective forms are summarized, called vector-to-vector and tensor-to-tensor forms, respectively.

However, for all machine learning algorithms, they have to train all samples again when new samples are added, which results in heavy computational cost. Consequently, incremental machine learning algorithms are proposed [13–17]. But most incremental learning algorithms focus on vector machine learning. Only a limited number of works study incremental learning in tensor space [18–20]. To investigate the incremental tensor learning, this paper develops incremental tensor discriminant analysis (ITDA), which employs supervised learning in tensor space and introduces incremental learning to process online learning. Furthermore, as a kind of machine learning algorithm, this paper also exploits the relationship between the proposed methods and the graph embedding framework and proves that the algorithm is a special case of tensor-to-tensor form under the graph embedding framework theoretically. This paper also analyzes the time and space complexity in detail. At last, this paper conducts facial image detection experiments to evaluate the proposed method. The experimental results have demonstrated the advantage of the method.

2. Tensor Discriminant Analysis

For multidimensional image data $X = \{X_1, \dots, X_K\}$, where $X_i \in \mathbb{R}^{I_1 \times \dots \times I_N}$, the corresponding class label is $l(i) \in [1, C]$, where C is the number of the class. Let the number of the c th class be n_c ; then the following definitions are introduced.

Definition 1. Within-class scatter tensor is defined:

$$S_w = \sum_{c=1}^C \sum_{i=1}^{n_c} \|X_i - \bar{X}_c\|^2, \quad (1)$$

where $\bar{X}_c$ represent the mean tensor of the c th class.

Definition 2. Between-class scatter tensor is defined:

$$S_b = \sum_{c=1}^C n_c \|\bar{X}_c - \bar{X}\|^2, \quad (2)$$

where $\bar{X}$ represent the total mean tensor.

Definition 3. Total scatter tensor is defined:

$$S_t = \sum_{i=1}^K \|X_i - \bar{X}\|^2. \quad (3)$$

It is easy to derive that

$$\begin{aligned} S_w + S_b &= \sum_{c=1}^C \sum_{i=1}^{n_c} \|X_i - \bar{X}_c\|^2 \\ &\quad + \sum_{c=1}^C n_c \|\bar{X}_c - \bar{X}\|^2 \end{aligned}$$

$$\begin{aligned} &= \sum_{c=1}^C \sum_{i=1}^{n_c} \|\text{vec}(X_i) - \text{vec}(\bar{X}_c)\|^2 \\ &\quad + \sum_{c=1}^C n_c \|\text{vec}(\bar{X}_c) - \text{vec}(\bar{X})\|^2 \\ &= \sum_{i=1}^K \|\text{vec}(X_i) - \text{vec}(\bar{X})\|^2 \\ &= S_t. \end{aligned} \quad (4)$$

Definition 4. Mode- n within-class scatter matrix is defined:

$$S_w^{(n)} = \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i^{(n)} - \bar{X}_c^{(n)}) (X_i^{(n)} - \bar{X}_c^{(n)})^T, \quad (5)$$

where $X_i^{(n)}$ is the mode- n matrix of the i th sample and $\bar{X}_c^{(n)}$ is the mode- n mean matrix of the c th class.

Definition 5. Mode- n between-class scatter matrix is defined:

$$S_b^{(n)} = \sum_{c=1}^C n_c (\bar{X}_c^{(n)} - \bar{X}^{(n)}) (\bar{X}_c^{(n)} - \bar{X}^{(n)})^T, \quad (6)$$

where $\bar{X}^{(n)}$ is the mode- n total mean matrix.

Definition 6. Mode- n total scatter matrix is defined:

$$\begin{aligned} S_t^{(n)} &= \sum_{i=1}^K (X_i^{(n)} - \bar{X}^{(n)}) (X_i^{(n)} - \bar{X}^{(n)})^T \\ &= S_b^{(n)} + S_w^{(n)}. \end{aligned} \quad (7)$$

The basic idea of TDA is to seek N projective matrices to make within-class scatter tensor smaller and between-class scatter tensor larger. The objective function is

$$\begin{aligned} J(U^{(1)}, \dots, U^{(N)}) &= \arg \max \frac{\sum_{c=1}^C n_c \|\bar{Y}_c - \bar{Y}\|^2}{\sum_{c=1}^C \sum_{i=1}^{n_c} \|Y_i - \bar{Y}_c\|^2} \\ &= \arg \max \frac{\sum_{c=1}^C n_c \left\| (\bar{X}_c - \bar{X}) \times_n U^{(n)T} \right\|_{n=1}^N^2}{\sum_{c=1}^C \sum_{i=1}^{n_c} \left\| (X_i - \bar{X}_c) \times_n U^{(n)T} \right\|_{n=1}^N^2}. \end{aligned} \quad (8)$$

In order to solve the above function, the iterative technique is adopted. It is assumed that the projective matrices

$\{U^{(1)}, \dots, U^{(n-1)}, U^{(n+1)}, \dots, U^{(N)}\}$ are known; then $U^{(n)}$ is solved as follows:

$$\begin{aligned}
 J(U^{(n)}) &= \arg \max \left(\left(\sum_{c=1}^C n_c U^{(n)T} (\bar{X}_c^{(n)} - \bar{X}^{(n)}) U^{(-n)} U^{(-n)T} \right. \right. \\
 &\quad \cdot (\bar{X}_c^{(n)} - \bar{X}^{(n)})^T U^{(n)} \Big) \\
 &\quad \cdot \left(\sum_{c=1}^C \sum_{i=1}^{n_c} U^{(n)T} (X_i^{(n)} - \bar{X}_c^{(n)}) U^{(-n)} U^{(-n)T} \right. \\
 &\quad \cdot (X_i^{(n)} - \bar{X}_c^{(n)})^T U^{(n)} \Big)^{-1} \Big), \tag{9}
 \end{aligned}$$

where $U^{(-n)} = U^{(N)} \dots \otimes U^{(n+1)} \otimes U^{(n-1)} \dots \otimes U^{(1)}$. Since $U^{(-n)} U^{(-n)T} = I$, so the above equation can be rewritten:

$$\begin{aligned}
 J(U^{(n)}) &= \arg \max \frac{\sum_{c=1}^C n_c U^{(n)T} (\bar{X}_c^{(n)} - \bar{X}^{(n)}) (\bar{X}_c^{(n)} - \bar{X}^{(n)})^T U^{(n)}}{\sum_{c=1}^C \sum_{i=1}^{n_c} U^{(n)T} (X_i^{(n)} - \bar{X}_c^{(n)}) (X_i^{(n)} - \bar{X}_c^{(n)})^T U^{(n)}} \\
 &= \arg \max \frac{U^{(n)T} S_b^{(n)} U^{(n)}}{U^{(n)T} S_w^{(n)} U^{(n)}}. \tag{10}
 \end{aligned}$$

Based on the basic concept of TDA and related matrix knowledge, we can get the following theorems.

Theorem 7. In tensor discriminant analysis, the mode- n intraclass scatter matrix is generally nonsingularity.

Proof. Defining the following matrix

$$H_w^{(n)} = [X_1^{(n)} - \bar{X}_{l(1)}^{(n)}, \dots, X_M^{(n)} - \bar{X}_{l(M)}^{(n)}], \tag{11}$$

where M is the number of samples, $l(m)$ expresses the class label of the m th sample. Then the mode- n intraclass scatter matrix is represented:

$$S_w^{(n)} = H_w^{(n)} H_w^{(n)T}, \tag{12}$$

where $H_w^{(n)} \in \mathbb{R}^{I_n \times D}$, $D = \prod_{k=1, k \neq n}^N I_k$. Generally speaking, $I_n \ll D$; then

$$\text{rank}(H_w^{(n)}) = \min(I_n, D) = I_n. \tag{13}$$

Further, we can get

$$\text{rank}(S_w^{(n)}) = \text{rank}(H_w^{(n)} H_w^{(n)T}) = \text{rank}(H_w^{(n)}) = I_n. \tag{14}$$

So the theorem is proved. $\square$

Theorem 8. Equation (8) can be unified into the graph embedding framework [12].

Proof. Based on the basic concept of tensor algebra, the numerator of (8) can be rewritten:

$$\begin{aligned}
 &\sum_{c=1}^C \sum_{i=1}^{n_c} \|Y_i - \bar{Y}_c\|^2 \\
 &= \sum_{c=1}^C \sum_{i=1}^{n_c} \|\text{vec}(Y_i) - \text{vec}(\bar{Y}_c)\|^2. \tag{15}
 \end{aligned}$$

Letting $y_i = \text{vec}(Y_i) \text{vec}(\bar{Y}_c) = (1/n_c) \sum_{j=1}^{n_c} y_j$, then the above equation is written:

$$\begin{aligned}
 &\sum_{c=1}^C \sum_{i=1}^{n_c} \left\| y_i - \frac{1}{n_c} \sum_{j=1}^{n_c} y_j \right\|^2 \\
 &= \sum_{c=1}^C \sum_{i=1}^{n_c} \left(y_i - \frac{1}{n_c} \sum_{j=1}^{n_c} y_j \right) \left(y_i - \frac{1}{n_c} \sum_{j=1}^{n_c} y_j \right)^T \\
 &= \sum_{i=1}^K y_i y_i^T - \sum_{c=1}^C \frac{1}{n_c} \sum_{l(i)=l(j)=c} y_i y_j^T \\
 &= \sum_{i=1}^K \sum_{j=1}^K W_{ij}^{(w)} y_i y_i^T - \sum_{i,j=1}^K W_{ij}^{(w)} y_i y_j^T \\
 &= \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K W_{ij}^{(w)} y_i y_i^T + \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K W_{ij}^{(w)} y_i y_i^T \\
 &\quad - \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(w)} y_i y_j^T \\
 &\quad - \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(w)} y_i y_j^T \\
 &= \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(w)} (y_i y_i^T + y_j y_j^T - y_i y_j^T - y_j y_i^T) \\
 &= \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(w)} (y_i - y_j)(y_i - y_j)^T \\
 &= \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(w)} \|Y_i - Y_j\|^2,
 \end{aligned} \tag{16}$$

where

$$W_{ij}^{(w)} = \begin{cases} \frac{1}{n_c}, & l(i) = l(j) = c, \\ 0, & l(i) \neq l(j). \end{cases} \tag{17}$$

Within the low-dimensional feature space, it is desired to preserve the property as demonstrated in (4), so the denominator of (8) is formulated as follows:

$$\begin{aligned}
& \sum_{c=1}^C n_c \|\bar{Y}_c - \bar{Y}\|^2 \\
&= \sum_{c=1}^C n_c \|\text{vec}(\bar{Y}_c) - \text{vec}(\bar{Y})\|^2 \\
&= \sum_{i=1}^K \|\text{vec}(\bar{Y}_i) - \text{vec}(\bar{Y})\|^2 \\
&\quad - \sum_{c=1}^C \sum_{i=1}^{n_c} \|\text{vec}(Y_i) - \text{vec}(\bar{Y}_c)\|^2 \\
&= \sum_{i=1}^K (y_i - \bar{y})(y_i - \bar{y})^T \\
&\quad - \sum_{c=1}^C \sum_{i=1}^{n_c} \|\text{vec}(Y_i) - \text{vec}(\bar{Y}_c)\|^2,
\end{aligned} \tag{18}$$

where

$$\begin{aligned}
& \sum_{i=1}^K (y_i - \bar{y})(y_i - \bar{y})^T \\
&= \sum_{i=1}^K y_i y_i^T - \frac{1}{K} \sum_{i,j=1}^K y_i y_j^T \\
&= \sum_{i=1}^K \left(\sum_{j=1}^K \frac{1}{K} \right) y_i y_i^T - \frac{1}{K} \sum_{i,j=1}^K y_i y_j^T \\
&= \frac{1}{2} \left(\frac{1}{K} \sum_{i,j=1}^K y_i y_i^T + \frac{1}{K} \sum_{i,j=1}^K y_i y_i^T - \frac{1}{K} \sum_{i,j=1}^K y_i y_j^T - \frac{1}{K} \sum_{i,j=1}^K y_j y_i^T \right) \\
&= \frac{1}{2} \sum_{i,j=1}^K \frac{1}{K} (y_i y_i^T + y_j y_j^T - y_i y_j^T - y_j y_i^T).
\end{aligned} \tag{19}$$

Combining (19) with (16), (18) can be written:

$$\begin{aligned}
& \sum_{c=1}^C n_c \|\bar{Y}_c - \bar{Y}\|^2 \\
&= \frac{1}{2} \sum_{i,j=1}^K \left(\frac{1}{K} - W_{ij}^{(w)} \right) (y_i y_i^T + y_j y_j^T - y_i y_j^T - y_j y_i^T) \\
&= \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(b)} (y_i y_i^T + y_j y_j^T - y_i y_j^T - y_j y_i^T)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(b)} (y_i - y_j)(y_i - y_j)^T \\
&= \frac{1}{2} \sum_{i,j=1}^K W_{ij}^{(b)} \|Y_i - Y_j\|^2,
\end{aligned} \tag{20}$$

where

$$W_{ij}^{(b)} = \frac{1}{n} - W_{ij}^{(w)} = \begin{cases} \frac{1}{K} - \frac{1}{n_c}, & l(i) = l(j) = c, \\ \frac{1}{K}, & l(i) \neq l(j). \end{cases} \tag{21}$$

Consequently, (8) is expressed:

$$\begin{aligned}
& J(U^{(1)}, \dots, U^{(N)}) \\
&= \arg \max \frac{\sum_{i,j=1}^K W_{ij}^{(b)} \|Y_i - Y_j\|^2}{\sum_{i,j=1}^K W_{ij}^{(w)} \|Y_i - Y_j\|^2} \\
&= \arg \max \frac{\sum_{i,j=1}^K W_{ij}^{(b)} \left\| (X_i - X_j) \times {}_n U^{(n)T} \Big|_{n=1}^N \right\|^2}{\sum_{i,j=1}^K W_{ij}^{(w)} \left\| (X_i - X_j) \times {}_n U^{(n)T} \Big|_{n=1}^N \right\|^2}.
\end{aligned} \tag{22}$$

The form of (22) is consistent with the tensor-to-tensor form of the graph embedding framework. Therefore, (8) can be unified into the graph embedding framework. $\square$

3. Incremental Tensor Discriminant Analysis

3.1. Incremental Learning Based on a Single Sample. In order to distinguish these variables that need to be updated during incremental learning procedure, the paper employs the subscript old to mark the variables before incremental learning. For example, $\bar{X}_{\text{old}}$ expresses the total mean tensor before new samples are added.

When a single sample X_{new} is added, its class label is l_{new} ; then the mode- n total mean matrix becomes

$$\bar{X}^{(n)} = \frac{K \bar{X}_{\text{old}}^{(n)} + X_{\text{new}}^{(n)}}{K + 1}. \tag{23}$$

If $l_{\text{new}} \notin [1, C]$, that is, the new sample belongs to a new class. In this case, the total class number is $C = C_{\text{old}} + 1$ and mode- n interclass scatter matrix is updated:

$$\begin{aligned}
S_b^{(n)} &= \sum_{c=1}^C n_c (\bar{X}_c^{(n)} - \bar{X}^{(n)}) (\bar{X}_c^{(n)} - \bar{X}^{(n)})^T \\
&= \sum_{c=1}^{C_{\text{old}}} n_{\text{old},c} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}^{(n)}) (\bar{X}_{\text{old},c}^{(n)} - \bar{X}^{(n)})^T \\
&\quad + (X_{\text{new}}^{(n)} - \bar{X}^{(n)}) (X_{\text{new}}^{(n)} - \bar{X}^{(n)})^T,
\end{aligned} \tag{24}$$

where n_c is the updated sample number of the c th class. Mode- n intraclass scatter matrix is

$$\begin{aligned}
 S_w^{(n)} &= \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i^{(n)} - \bar{X}_c^{(n)}) (X_i^{(n)} - \bar{X}_c^{(n)})^T \\
 &= \sum_{c=1}^{C_{old}+1} \sum_{i=1}^{n_c} (X_i^{(n)} - \bar{X}_c^{(n)}) (X_i^{(n)} - \bar{X}_c^{(n)})^T \\
 &= \sum_{c=1}^{C_{old}} \sum_{i=1}^{n_{old,c}} (X_i^{(n)} - \bar{X}_{old,c}^{(n)}) (X_i^{(n)} - \bar{X}_{old,c}^{(n)})^T \\
 &\quad + (X_{new}^{(n)} - \bar{X}_{new}^{(n)}) (X_{new}^{(n)} - \bar{X}_{new}^{(n)})^T,
 \end{aligned} \tag{25}$$

where $\bar{X}_{new}^{(n)}$ is the mode- n mean matrix of the new sample. Because a single sample is added and it belongs to a new class, we can get

$$X_{new}^{(n)} = \bar{X}_{new}^{(n)}. \tag{26}$$

Then (25) becomes

$$S_w^{(n)} = \sum_{c=1}^{C_{old}} \sum_{i=1}^{n_{old,c}} (X_i^{(n)} - \bar{X}_{old,c}^{(n)}) (X_i^{(n)} - \bar{X}_{old,c}^{(n)})^T = S_{old,w}^{(n)}. \tag{27}$$

It is demonstrated in (27) that mode- n intraclass scatter matrix will not change when a new sample with new class is added.

When the class label of the new sample $l_{new} = r \in [1, C_{old}]$, that is, the class label is not a new class. In this case, the total class number $C = C_{old}$; then mode- n interclass scatter matrix is

$$S_b^{(n)} = \sum_{c=1}^C n_c (\bar{X}_c^{(n)} - \bar{X}^{(n)}) (\bar{X}_c^{(n)} - \bar{X}^{(n)})^T. \tag{28}$$

Mode- n intraclass scatter matrix is

$$\begin{aligned}
 S_w^{(n)} &= \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i^{(n)} - \bar{X}_c^{(n)}) (X_i^{(n)} - \bar{X}_c^{(n)})^T \\
 &= \sum_{c=1}^{C_{old}+1} \sum_{i=1}^{n_c} (X_i^{(n)} - \bar{X}_c^{(n)}) (X_i^{(n)} - \bar{X}_c^{(n)})^T \\
 &= \sum_{c=1}^{C_{old}} \sum_{i=1}^{n_{old,c}} (X_i^{(n)} - \bar{X}_{old,c}^{(n)}) (X_i^{(n)} - \bar{X}_{old,c}^{(n)})^T \\
 &\quad + \sum_{k=1}^{n_{old,r}+1} (X_k^{(n)} - \bar{X}_r^{(n)}) (X_k^{(n)} - \bar{X}_r^{(n)})^T.
 \end{aligned} \tag{29}$$

Because the new sample belongs to the r th class, then the class mean of the r th class becomes

$$\bar{X}_r^{(n)} = \frac{n_{old,r} \bar{X}_{old,r}^{(n)} + X_{new}^{(n)}}{n_{old,r} + 1}. \tag{30}$$

Based on this, we can get

$$\begin{aligned}
 &\sum_{k=1}^{n_{old,r}+1} (X_k^{(n)} - \bar{X}_r^{(n)}) (X_k^{(n)} - \bar{X}_r^{(n)})^T \\
 &= \sum_{k=1}^{n_{old,r}} (X_k^{(n)} - \bar{X}_r^{(n)}) (X_k^{(n)} - \bar{X}_r^{(n)})^T \\
 &\quad + (X_{new}^{(n)} - \bar{X}_r^{(n)}) (X_{new}^{(n)} - \bar{X}_r^{(n)})^T \\
 &= \sum_{k=1}^{n_{old,r}} (X_k^{(n)} - \bar{X}_{old,r}^{(n)} + \bar{X}_{old,r}^{(n)} - \bar{X}_r^{(n)}) \\
 &\quad \cdot (X_k^{(n)} - \bar{X}_{old,r}^{(n)} + \bar{X}_{old,r}^{(n)} - \bar{X}_r^{(n)})^T \\
 &\quad + \left(X_{new}^{(n)} - \bar{X}_{old,r}^{(n)} - \frac{X_{new}^{(n)} - \bar{X}_{old,r}^{(n)}}{n_{old,r} + 1} \right) \\
 &\quad \cdot \left(X_{new}^{(n)} - \bar{X}_{old,r}^{(n)} - \frac{X_{new}^{(n)} - \bar{X}_{old,r}^{(n)}}{n_{old,r} + 1} \right)^T \\
 &= \sum_{k=1}^{n_{old,r}} (X_k^{(n)} - \bar{X}_{old,r}^{(n)}) (X_k^{(n)} - \bar{X}_{old,r}^{(n)})^T \\
 &\quad + \frac{n_{old,r}}{n_{old,r} + 1} (X_{new}^{(n)} - \bar{X}_{old,r}^{(n)}) (X_{new}^{(n)} - \bar{X}_{old,r}^{(n)})^T.
 \end{aligned} \tag{31}$$

So (29) is simplified:

$$\begin{aligned}
 S_w^{(n)} &= \sum_{c=1}^{C_{old}} \sum_{i=1}^{n_{old,c}} (X_i^{(n)} - \bar{X}_{old,c}^{(n)}) (X_i^{(n)} - \bar{X}_{old,c}^{(n)})^T \\
 &\quad + \sum_{k=1}^{n_{old,r}} (X_k^{(n)} - \bar{X}_{old,r}^{(n)}) (X_k^{(n)} - \bar{X}_{old,r}^{(n)})^T \\
 &\quad + \frac{n_{old,r}}{n_{old,r} + 1} (X_{new}^{(n)} - \bar{X}_{old,r}^{(n)}) (X_{new}^{(n)} - \bar{X}_{old,r}^{(n)})^T \\
 &= \sum_{c=1}^{C_{old}} \sum_{i=1}^{n_{old,c}} (X_i^{(n)} - \bar{X}_{old,c}^{(n)}) (X_i^{(n)} - \bar{X}_{old,c}^{(n)})^T \\
 &\quad + \frac{n_{old,r}}{n_{old,r} + 1} (X_{new}^{(n)} - \bar{X}_{old,r}^{(n)}) (X_{new}^{(n)} - \bar{X}_{old,r}^{(n)})^T.
 \end{aligned} \tag{32}$$

3.2. Incremental Learning Based on Multisamples. When several samples are added, new added samples $X_{new} = \{X_{K+1}, \dots, X_{K+T}\}$, $T \geq 1$, the corresponding class labels are $l_{new} = \{l_1, \dots, l_T\}$. Without loss of generality, it is assumed that

$n_{\text{new},r}$ samples belong to the r th class; then the mean tensor of the r th class is updated:

$$\begin{aligned}\bar{X}_r &= \frac{1}{n_{\text{old},r} + n_{\text{new},r}} \left(n_{\text{old},r} \bar{X}_{\text{old},r} + \sum_{i=1}^{n_{\text{new},r}} X_i \right) \\ &= \frac{(n_{\text{old},r} \bar{X}_{\text{old},r} + n_{\text{new},r} \bar{X}_{\text{new},r})}{n_{\text{old},r} + n_{\text{new},r}},\end{aligned}\quad (33)$$

where $\bar{X}_{\text{new},r}$ is the mean tensor of the new samples belonging to the r th class. The corresponding mode- n mean matrix of the r th class is

$$\bar{X}_r^{(n)} = \frac{(n_{\text{old},r} \bar{X}_{\text{old},r}^{(n)} + n_{\text{new},r} \bar{X}_{\text{new},r}^{(n)})}{n_{\text{old},r} + n_{\text{new},r}}. \quad (34)$$

Then the number of samples in the r th is

$$n_r = n_{\text{old},r} + n_{\text{new},r}. \quad (35)$$

The total mean tensor is updated:

$$\bar{X} = \frac{K \bar{X}_{\text{old}} + \sum_{i=K+1}^{K+T} X_i}{K + T} = \frac{K \bar{X}_{\text{old}} + T \bar{X}_{\text{new}}}{K + T}, \quad (36)$$

where $\bar{X}_{\text{new}}$ is the mean tensor of all new samples. The interclass scatter mean tensor is updated:

$$S_b = \sum_{c=1}^C n_c \|\bar{X}_c - \bar{X}\|^2. \quad (37)$$

The corresponding mode- n interclass scatter matrix is

$$S_b^{(n)} = \sum_{c=1}^C n_c (\bar{X}_c^{(n)} - \bar{X}^{(n)}) (\bar{X}_c^{(n)} - \bar{X}^{(n)})^T. \quad (38)$$

The mode- n intraclass scatter matrix is

$$\begin{aligned}S_w^{(n)} &= \sum_{c=1}^C \sum_{i=1}^{n_c} (X_i^{(n)} - \bar{X}_c^{(n)}) (X_i^{(n)} - \bar{X}_c^{(n)})^T \\ &= \sum_{c=1}^C \left(\sum_{i=1}^{n_{\text{old},c}} (X_i^{(n)} - \bar{X}_{\text{old},c}^{(n)}) (X_i^{(n)} - \bar{X}_{\text{old},c}^{(n)})^T \right. \\ &\quad \left. + n_{\text{old},c} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_c^{(n)}) (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_c^{(n)})^T \right) \\ &\quad + \sum_{c=1}^C \left(\sum_{i=1}^{n_{\text{new},c}} (X_i^{(n)} - \bar{X}_{\text{new},c}^{(n)}) (X_i^{(n)} - \bar{X}_{\text{new},c}^{(n)})^T \right. \\ &\quad \left. + n_{\text{new},c} (\bar{X}_{\text{new},c}^{(n)} - \bar{X}_c^{(n)}) (\bar{X}_{\text{new},c}^{(n)} - \bar{X}_c^{(n)})^T \right). \quad (39)\end{aligned}$$

Substituting (34) into the following equation, we can get

$$\begin{aligned}&n_{\text{old},c} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_c^{(n)}) (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_c^{(n)})^T \\ &= n_{\text{old},c} \left(\bar{X}_{\text{old},c}^{(n)} - \frac{n_{\text{old},c} \bar{X}_{\text{old},c}^{(n)} + n_{\text{new},c} \bar{X}_{\text{new},c}^{(n)}}{n_{\text{old},c} + n_{\text{new},c}} \right) \\ &\quad \cdot \left(\bar{X}_{\text{old},c}^{(n)} - \frac{n_{\text{old},c} \bar{X}_{\text{old},c}^{(n)} + n_{\text{new},c} \bar{X}_{\text{new},c}^{(n)}}{n_{\text{old},c} + n_{\text{new},c}} \right)^T \\ &= \frac{n_{\text{old},c} n_{\text{new},c}^2}{(n_{\text{old},c} + n_{\text{new},c})^2} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)}) (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)})^T \\ &= \frac{n_{\text{old},c} n_{\text{new},c}^2}{n_c^2} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)}) (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)})^T.\end{aligned}\quad (40)$$

Similarly, we can get

$$\begin{aligned}&n_{\text{new},c} (\bar{X}_{\text{new},c}^{(n)} - \bar{X}_c^{(n)}) (\bar{X}_{\text{new},c}^{(n)} - \bar{X}_c^{(n)})^T \\ &= \frac{n_{\text{new},c} n_{\text{old},c}^2}{(n_{\text{old},c} + n_{\text{new},c})^2} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)}) \\ &\quad \cdot (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)})^T \\ &= \frac{n_{\text{new},c} n_{\text{old},c}^2}{n_c^2} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)}) \\ &\quad \cdot (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)})^T.\end{aligned}\quad (41)$$

Substituting (40) and (41) into (39), we can obtain

$$\begin{aligned}S_w^{(n)} &= \sum_{c=1}^n \left(\sum_{i=1}^{n_{\text{old},c}} (X_i^{(n)} - \bar{X}_{\text{old},c}^{(n)}) (X_i^{(n)} - \bar{X}_{\text{old},c}^{(n)})^T \right. \\ &\quad \left. + n_{\text{old},c} (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_c^{(n)}) (\bar{X}_{\text{old},c}^{(n)} - \bar{X}_c^{(n)})^T \right) \\ &\quad + \sum_{c=1}^n \left(\sum_{i=1}^{n_{\text{new},c}} (X_i^{(n)} - \bar{X}_{\text{new},c}^{(n)}) (X_i^{(n)} - \bar{X}_{\text{new},c}^{(n)})^T \right. \\ &\quad \left. + n_{\text{new},c} (\bar{X}_{\text{new},c}^{(n)} - \bar{X}_c^{(n)}) (\bar{X}_{\text{new},c}^{(n)} - \bar{X}_c^{(n)})^T \right) \\ &= \sum_{c=1}^n \left(\sum_{i=1}^{n_{\text{old},c}} (X_i^{(n)} - \bar{X}_{\text{old},c}^{(n)}) (X_i^{(n)} - \bar{X}_{\text{old},c}^{(n)})^T \right. \\ &\quad \left. + \sum_{i=1}^{n_{\text{new},c}} (X_i^{(n)} - \bar{X}_{\text{new},c}^{(n)}) (X_i^{(n)} - \bar{X}_{\text{new},c}^{(n)})^T \right)\end{aligned}$$

$$\begin{aligned}
& + \sum_{c=1}^n \frac{n_{\text{old},c} n_{\text{new},c}}{n_c} \left(\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)} \right) \\
& \quad \cdot \left(\bar{X}_{\text{old},c}^{(n)} - \bar{X}_{\text{new},c}^{(n)} \right)^T.
\end{aligned} \quad (42)$$

Without loss of generality, it is supposed that, for T new samples, there are n_{C+1} samples belonging to the new class label $C+1$; then updated mode- n interclass scatter matrix is

$$\begin{aligned}
S_b^{(n)} &= \sum_{c=1}^{C_{\text{old}}} n_c \left(\bar{X}_c^{(n)} - \bar{X}^{(n)} \right) \left(\bar{X}_c^{(n)} - \bar{X}^{(n)} \right)^T \\
&+ n_{C+1} \left(\bar{X}_{C+1}^{(n)} - \bar{X}^{(n)} \right) \left(\bar{X}_{C+1}^{(n)} - \bar{X}^{(n)} \right)^T \\
&= \sum_{c=1}^{C+1} n_c \left(\bar{X}_c^{(n)} - \bar{X}^{(n)} \right) \left(\bar{X}_c^{(n)} - \bar{X}^{(n)} \right)^T
\end{aligned} \quad (43)$$

and mode- n intraclass scatter matrix is

$$\begin{aligned}
S_w^{(n)} &= \sum_{c=1}^{C_{\text{old}}} \sum_{i=1}^{n_c} \left(X_i^{(n)} - \bar{X}_c^{(n)} \right) \left(X_i^{(n)} - \bar{X}_c^{(n)} \right)^T \\
&+ \sum_{i=1}^{n_{C+1}} \left(X_i^{(n)} - \bar{X}_{C+1}^{(n)} \right) \left(X_i^{(n)} - \bar{X}_{C+1}^{(n)} \right)^T.
\end{aligned} \quad (44)$$

It is not difficult to find that incremental learning based on singular sample only is a special case of incremental learning based on multisample.

3.3. The Complexity Analysis. For tensor discriminant analysis, the main computational time is spent on the computation of interclass mean, total mean, inter- and intraclass scatter tensor, and Eigen decomposition. The computation cost of inter- and intra-class scatter tensors depends on the number of training samples. If there are a large number of training samples, it cannot avoid to increment computational time.

For incremental discriminant analysis, the main computational time is spent on the computation of updated inter- and intraclass scatter matrix and the class number.

For Eigen decomposition, both the time complexity of TDA and ITDA are $O(NI^3)$. The main difference of the time complexity is the computation of inter- and intraclass scatter matrix. For TDA, the time complexity is $O(MNI^{N+1})$, so the time complexity will increase with the number of training samples. For ITDA, the time complexity is $O(TNI^{N+1} + CNI^{N+1})$, which is related to the class number and the number of new samples. It has no relationship with the number of initial training samples. Consequently, ITDA is helpful to reduce the time complexity.

Considering the space complexity, ITDA is also superior to TDA. When new samples are added, TDA needs $M \prod_{n=1}^N I_n$ bytes to save all training samples, but ITDA only needs $\prod_{n=1}^N I_n$ bytes to save new added samples, $\prod_{n=1}^N I_n$ bytes to save the total mean, $C \prod_{n=1}^N I_n$ bytes to save the class mean, and

FIGURE 1: The samples of CBCL dataset.

FIGURE 2: The detection results after the first incremental learning.

$N \sum_{n=1}^N I_n^2$ bytes to save mode- n scatter matrix. Hence ITDA has the capability to save space.

Compared to incremental learning based on single sample with incremental learning based on multisamples, incremental learning based on single samples has an advantage to reduce the space complexity because it only deals with one sample for each time.

4. Experiments

In this section, a series of experiments are carried out to validate the performance of incremental tensor discriminant analysis (ITDA). The CBCL image data set is used to conduct facial image detection experiments. The dataset contains two classes of images, including facial images and nonfacial images as shown in Figure 1. The total number of the datasets is 2988 images, in which there are 2429 facial images and 559 nonfacial images. For each image, the size is 19×19 . This paper divides whole dataset into training dataset with 1215 facial images and 280 nonfacial images and testing dataset with 1214 facial images and 279 nonfacial images. Furthermore, training dataset is divided into initial training

FIGURE 3: The detection results after the second incremental learning.

dataset with 1015 facial images and 80 nonfacial images and four incremental datasets. Each incremental dataset has 50 facial images and 50 nonfacial images.

ITLDA integrates the tensor representation and incremental learning; it is reasonable to believe that it has the advantage to improve the detection performance and reduce the time and space complexity. In this respect, ITLDA is compared with LDA [21], ILDA [14], TPCA [22], ITPCA [23], and TDA [9]. LDA is the classical linear discriminant analysis. ILDA is the incremental version of LDA. TPCA is also called MPCA (multilinear principal component analysis), which carries on principal component analysis with tensor data. ITPCA is proposed to suit for incremental principal component analysis for tensor data. TDA also represents data as tensor structure and conducts multilinear discriminant analysis. For each time of incremental learning, the paper adds one incremental dataset and then extracts low-dimensional features on testing dataset. The nearest neighbor classifier is employed to classify these low-dimensional features.

The comparisons of detection performance for different algorithms with incremental learning are shown in Figures 2, 3, 4, and 5, respectively. It is worth noting that LDA is the worst and ILDA is better than LDA. However the detection results of ILDA drop with the increment of the dimension of low-dimensional features. TPCA and ITPCA have similar detection results and both of them exceed LDA and ILDA. The probable reason is that TPCA and ITPCA represent data as tensor structure, which make full use of the interior structure information to enhance the detection performance. TDA is superior to the above four algorithms. When the dimension of low-dimensional features is low, TDA and ITDA have comparative detection percent and ITDA begins to surmount TDA when the dimension of low-dimensional features increases. Figure 6 and Table 1 have

FIGURE 4: The detection results after the third incremental learning.

FIGURE 5: The detection results after the fourth incremental learning.

shown the best detection results of different algorithms. It can be seen that the detection performances of different algorithms are improved with the increment of incremental learning numbers and ITLDA always has the best performance. Consequently, it can be derived that the increment of incremental learning number is helpful to improve the detection result. More than that, as shown in Figures 7 and 8, incremental learning algorithms ILDA, ITPCA, and ITDA have the capability to alleviate time and space complexity

TABLE 1: The best detection results of different algorithms with incremental learning.

Algorithms	The first incremental learning (%)	The second incremental learning (%)	The third incremental learning (%)	The fourth incremental learning (%)
LDA	82.59	81.51	82.12	82.52
ILDA	90.56	91.49	91.56	91.56
TPCA	90.22	91.36	92.03	92.83
ITPCA	88.55	90.56	91.63	92.1
TLDA	93.67	94.44	95.38	95.78
ITLDA	93.83	95.17	96.05	96.45

FIGURE 6: The comparison of the best detection results for different algorithms with incremental learning.

FIGURE 8: The comparison of space complexity.

FIGURE 7: The comparison of time complexity.

apparently compared with nonincremental learning. Furthermore, since ITPCA and ITDA adopt tensor representation, they have lower time and space requirements than LDA.

5. Conclusions

In this paper, incremental tensor discriminant analysis (ITDA) is investigated. It adopts tensor representation to keep the structure information for high-dimensional images and introduces incremental learning to complete online learning. This paper also proves the relationship between ITDA and the graph framework theoretically. The facial detection experiments have shown that ITDA has better performance than TDA and is able to reduce the time and space complexity apparently.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This present work has been funded with support from the Young Scientist Project of Chengdu University (no. 2013XJZ21), Project of Science and Technology Support Program of Sichuan Province, China, under Grant no. 2014GZ0013, and Project of Education Department of Sichuan Province, China, under Grant no. 13ZA0297.

References

- [1] M. Murtaza, M. Sharif, M. Raza, and J. Shah, "Face recognition using adaptive margin fisher's criterion and linear discriminant analysis," *International Arab Journal of Information Technology*, vol. 11, no. 2, pp. 1–11, 2014.
- [2] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 10, no. 2, pp. 159–203, 1948.
- [3] C. Liu, K. He, J.-L. Zhou, and C.-B. Gao, "Discriminant orthogonal rank-one tensor projections for face recognition," in *Intelligent Information and Database Systems*, vol. 6592 of *Lecture Notes in Computer Science*, pp. 203–211, Springer, Berlin, Germany, 2011.
- [4] G.-F. Lu, Z. Lin, and Z. Jin, "Face recognition using discriminant locality preserving projections based on maximum margin criterion," *Pattern Recognition*, vol. 43, no. 10, pp. 3572–3579, 2010.
- [5] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 103–123, 2009.
- [6] J.-L. Minoi, C. E. Thomaz, and D. F. Gillies, "Tensor-based multivariate statistical discriminant methods for face applications," in *Proceedings of the International Conference on Statistics in Science, Business, and Engineering (ICSSBE '12)*, pp. 1–6, IEEE, Langkawi, Malaysia, September 2012.
- [7] F. Nie, S. Xiang, Y. Song, and C. Zhang, "Extracting the optimal dimensionality for local tensor discriminant analysis," *Pattern Recognition*, vol. 42, no. 1, pp. 105–114, 2009.
- [8] N. Tang, X. Z. Gao, and X. Li, "Tensor subclass discriminant analysis for radar target classification," *Electronics Letters*, vol. 48, no. 8, pp. 455–456, 2012.
- [9] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.
- [10] S.-J. Wang, J. Yang, M.-F. Sun, X.-J. Peng, M.-M. Sun, and C.-G. Zhou, "Sparse tensor discriminant color space for face verification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 6, pp. 876–888, 2012.
- [11] Z.-Z. Yu, C.-C. Jia, W. Pang, C.-Y. Zhang, and L.-H. Zhong, "Tensor discriminant analysis with multiscale features for action modeling and categorization," *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 95–98, 2012.
- [12] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.
- [13] A. Joseph, Y.-M. Jang, S. Ozawa, and M. Lee, "Extension of incremental linear discriminant analysis to online feature extraction under nonstationary environments," in *Neural Information Processing*, vol. 7664 of *Lecture Notes in Computer Science*, pp. 640–647, Springer, Berlin, Germany, 2012.
- [14] G.-F. Lu, J. Zou, and Y. Wang, "Incremental complete LDA for face recognition," *Pattern Recognition*, vol. 45, no. 7, pp. 2510–2521, 2012.
- [15] G.-F. Lu, J. Zou, and Y. Wang, "Incremental learning of complete linear discriminant analysis for face recognition," *Knowledge-Based Systems*, vol. 31, pp. 19–27, 2012.
- [16] G.-F. Lu, J. Zou, and Y. Wang, "Incremental learning of discriminant common vectors for feature extraction," *Applied Mathematics and Computation*, vol. 218, no. 22, pp. 11269–11278, 2012.
- [17] Q. Wang and L. Zhang, "Least squares online linear discriminant analysis," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1510–1517, 2012.
- [18] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos, "Incremental tensor analysis: theory and applications," *ACM Transactions on Knowledge Discovery from Data*, vol. 2, no. 3, article 11, 2008.
- [19] J.-G. Wang, E. Sung, and W.-Y. Yau, "Incremental two-dimensional linear discriminant analysis with applications to face recognition," *Journal of Network and Computer Applications*, vol. 33, no. 3, pp. 314–322, 2010.
- [20] J. Wen, X. Gao, Y. Yuan, D. Tao, and J. Li, "Incremental tensor biased discriminant analysis: a new color-based visual tracking method," *Neurocomputing*, vol. 73, no. 4–6, pp. 827–839, 2010.
- [21] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [22] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.
- [23] C. Liu, T. Yan, W. Zhao et al., "Incremental tensor principal component analysis for handwritten digit recognition," *Mathematical Problems in Engineering*, vol. 2014, Article ID 819758, 10 pages, 2014.