Technical Correspondence

Incremental Embedding and Learning in the Local Discriminant Subspace With Application to Face Recognition

Miao Cheng, *Student Member, IEEE*, Bin Fang, *Member, IEEE*, Yuan Yan Tang, *Fellow, IEEE*, Taiping Zhang, *Student Member, IEEE*, and Jing Wen, *Student Member, IEEE*

***Abstract*—Dimensionality reduction and incremental learning have recently received broad attention in many applications of data mining, pattern recognition, and information retrieval. Inspired by the concept of manifold learning, many discriminant embedding techniques have been introduced to seek low-dimensional discriminative manifold structure in the high-dimensional space for feature reduction and classification. However, such graph-embedding framework-based subspace methods usually confront two limitations: 1) since there is no available updating rule for local discriminant analysis with the additive data, it is difficult to design incremental learning algorithm and 2) the small sample size (SSS) problem usually occurs if the original data exist in very high-dimensional space. To overcome these problems, this paper devises a supervised learning method, called local discriminant subspace embedding (LDSE), to extract discriminative features. Then, the incremental-mode algorithm, incremental LDSE (ILDSE), is proposed to learn the local discriminant subspace with the newly inserted data, which applies incremental learning extension to the batch LDSE algorithm by employing the idea of singular value-decomposition (SVD) updating algorithm. Furthermore, the SSS problem is avoided in our method for the high-dimensional data and the benchmark incremental learning experiments on face recognition show that ILDSE bears much less computational cost compared with the batch algorithm.**

***Index Terms*—Dimensionality reduction, discriminant embedding, face recognition, incremental learning, manifold learning, singular value decomposition (SVD), small sample size (SSS) problem.**

I. Introduction

It is well known that the problem of feature extraction is one of the central issues for data mining and classification. Efficient feature-extraction method can improve the classification results in the reduced subspace. In the past decades, numerous dimension reduction methods have been proposed to find the low-dimensional feature representation. The most popular techniques for this purpose may be principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2]. Rather than PCA, LDA [2], [3] is a supervised feature-extraction technique for pattern recognition, which tends to find a set of projective direction to maximize the between-class distance and minimize the within-class distance simultaneously. An intrinsic limitation of LDA is that it usually suffers from the *small sample size* (SSS) problem, where the sample size is much smaller than the size of dimensionality of samples. To address this problem, many efficient LDA algorithms, e.g., PCA + LDA [3], direct LDA (DLDA) [4], LDA/QR [5], have been developed to handle the high-dimensional data and learn the discriminant subspace. In addition, the model-based learning algorithms, e.g., geometric mean [6] and harmonic mean [7] based subspace selection [8], associated with some convention model-based algorithms [9], can improve the discriminant performance of LDA.

Different from the global dimensionality reduction methods, manifold learning aims to find the low-dimensional manifold structure embedded in the input space. The theoretical foundation of manifold learning is based on the consideration that the high-dimensional data may lie on an intrinsic nonlinear manifold, and the dimensionality of which is much lower. The most representative manifold learning algorithms include ISOMAP [10], locally linear embedding (LLE) [11], self-organizing maps (SOM) [12], and Laplacian eigenmaps [13]–[15]. However, these methods are designed to preserve the local geometrical structure of original high-dimensional data in the lower dimensional space rather than good discriminant ability.

In order to exploit the local discriminative manifold structure, some supervised learning techniques are proposed by incorporating the class information into the locality-preserving learning techniques [16]–[18]. Recently, Yan *et al.* [19] explains the manifold learning techniques and a large number of popular dimensionality reduction methods as a general framework that can be defined in a graph-embedding way instead of a kernel view [20] corresponding to the formulation of kernel PCA [21]. Moreover, some recently proposed patch-based algorithm [22], [23] can also be absorbable into the graph-embedding approach, with respect to the alignment framework [24]. Generally, the discriminative feature-extraction algorithms are summarized as a graph-based constraint embedding by defining the intrinsic and penalty graphs. In other words, it finds a set of projection directions in the linear embedded subspace, i.e.,

$$J(w) = \arg \min_{wXBX^T w = c} w^T XLX^T w = \arg\min_{w} \frac{w^T XLX^T w}{w^T XBX^T w} \quad (1)$$

where c is a constant, X is the data matrix, L is the Laplacian matrix of intrinsic graph, which is defined as follows:

$$L = D - S, \qquad D_{ii} = \sum_j S_{ij}. \quad (2)$$

Here, S is the affinity matrix of the intrinsic graph and D is a diagonal matrix, of which diagonal entries are column (or row) sum of S. In addition, B can be a diagonal matrix that is used for scale normalization, or the Laplacian matrix of a penalty graph G', $L_p = D_p - S_p$, where S_p indicates the adjacency matrix of penalty graph that describes the similarity of extraclass data that should be avoided for classification and D_p is the diagonal matrix defined in graph-embedding framework. For the general discriminant embedding, one solution of the optimal w is the eigenvector corresponding to the smallest eigenvalue of the generalized eigenvalue decomposition (ED) $XLX^T w = \lambda XBX^T w$ that is usually computed via eigenanalysis on $(XBX^T)^{-1}(XLX^T)$, for more tractable.

In the recent years, incremental learning has already attracted much attention as a result of the increasing demand for developing machine vision/intelligent systems, and a dozen of incremental learning algorithms have been proposed, especially in the data-mining domain [25]–[27] and the image-retrieval field [28]. Until now, several efficient incremental PCA algorithms have been developed, as well as few techniques that deal with the incremental learning for LDA, due

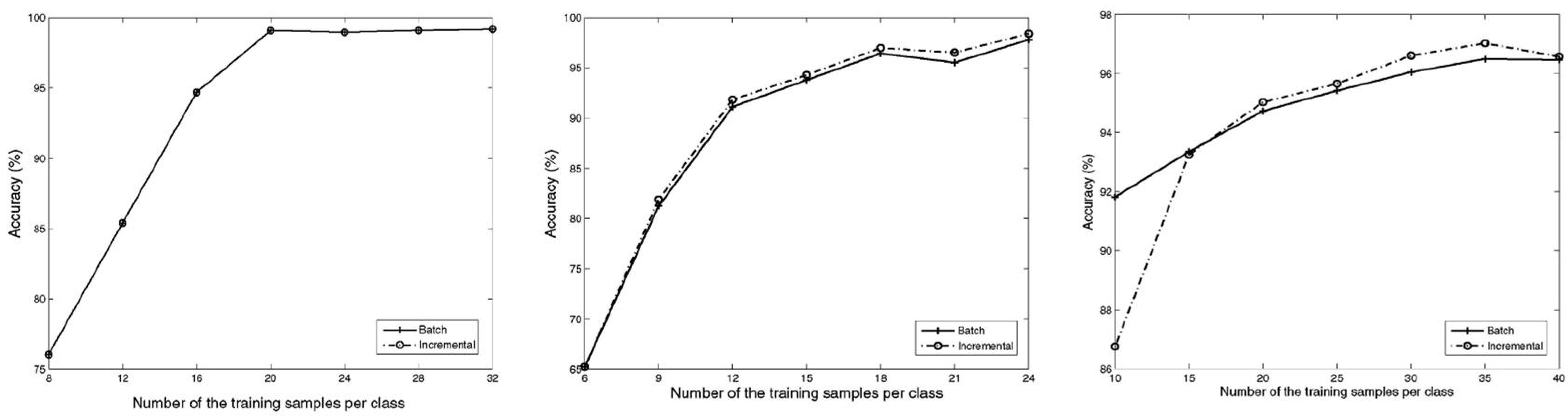
Manuscript received April 21, 2009; revised July 14, 2009 and November 19, 2009; accepted February 7, 2010. Date of publication March 22, 2010; date of current version August 18, 2010. This work was supported by the New Century Excellent Talents Program of Educational Ministry of China (NCET-06-0762), by the Natural Science Foundation of China under Grant NSF-60873092 and 90820306, and by the Natural Science Foundation of Chongqing under Grant CSTC2007BA2003. This paper was recommended by Associate Editor M. Last.

The authors are with the Department of Computer Science, Chongqing University, Chongqing 400030, China (e-mail: chengm@cqu.edu.cn; fb@cqu.edu.cn; yytang@cqu.edu.cn; tpzhang@cqu.edu.cn; wj@cqu.edu.cn).

Digital Object Identifier 10.1109/TSMCC.2010.2043529

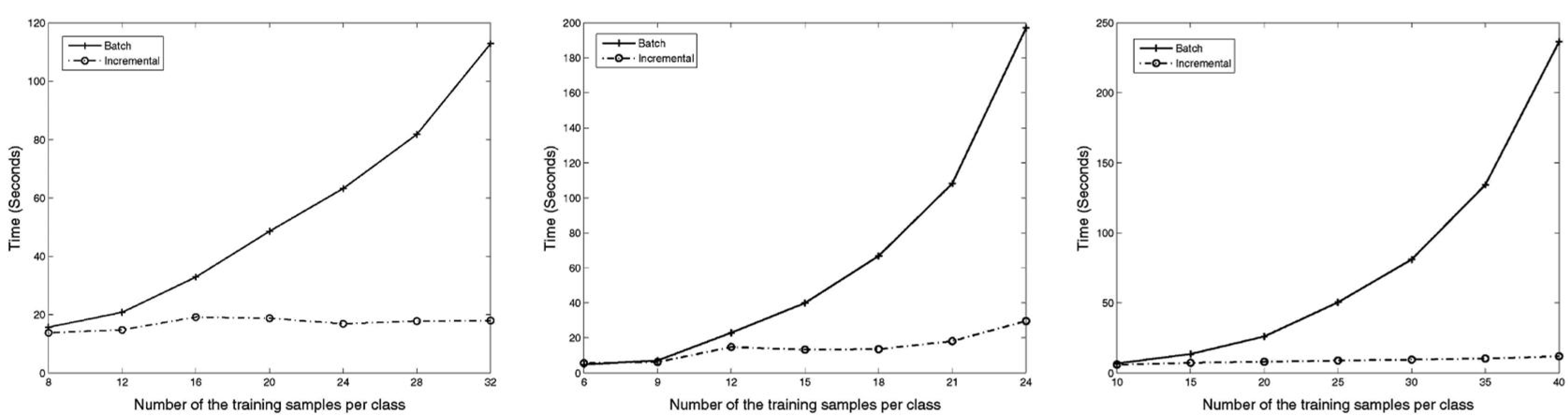
TABLE II
STATISTICAL COMPARISONS OF DIFFERENT ALGORITHMS

Measurement	PCA	LPP	SLPP	PCA+LDA	LPP+LDA	DLDA	LDA/QR	MFA	LDSE
Rank	8.46667	6.66667	5.6	2.46667	8.53333	4.66667	3.26667	3.66667	1.66667
Rank^2	71.6845	44.44449	31.36	6.08446	72.81772	21.77781	10.67113	13.44447	2.77779
Difference	6.8	5	3.93333	0.8	6.86666	3	1.6	2	–



(a) (b) (c)

Fig. 8. Comparison of classification accuracy rate between batch LDSE and incremental LDSE with the evenly additive samples. (a) CMU PIE (b) AR. (c) EYale B.



(a) (b) (c)

Fig. 9. Comparison of computational time between batch LDSE and incremental LDSE with the evenly additive samples. (a) CMU PIE. (b) AR. (c) EYale B.

C. Performance of Incremental LDSE

In this section, we evaluate the performance of incremental LDSE for the newly inserted data and three respective experiments are performed. At the beginning of experiments, for each individual in the CMU PIE, AR, and EYale B face databases, the images are sorted in a random order. Likewise, the experiments are repeated ten times with different training images, the average accuracies and the average spending time are recorded, where the KNN classifier with $K = 1$ is employed for classification.

1) Incremental Embedding With the Evenly Additive Samples: In the CMU PIE database, initially, we select the first four random images of each person for training using batch LDSE. For each time, four images per person are inserted into the training set incrementally with the given order, and the remaining data are used as the testing set. For each individual in the AR database, the first three random images are selected for training by employing batch LDSE and another three images are added for incremental learning. In the EYale B database, the first five images from each subject are selected for training in batch mode, five other images are added for training at each learning time. The parametric pair $(\varepsilon_1, \varepsilon_2)$ is set to be (100, 20), (200, 45), and (80, 35) for CMU PIE, AR, and EYale B databases, respectively. The experimental results are shown in Figs. 8 and 9.

2) Incremental Embedding With the Unevenly Additive Samples: In the three face databases, about 10% images from each individual (4 images $\times$ 68 persons $=$ 272 images for CMU PIE, 3 images $\times$ 50 persons $=$ 150 images for AR, and 5 images $\times$ 38 subjects $=$ 190 images for EYale B) are selected for training using batch LDSE. Then, all the remaining images are sorted in a random order. For each time, about 10% images (292 images in CMU PIE, 130 images in AR, and 171 images in EYale B) are inserted into the training set for incremental learning by performing ILDSE, and the surplus images are used for testing. The setting of $(\varepsilon_1, \varepsilon_2)$ for the three face databases follows the initialization in the previous experiment. The incremental discriminant results and computational time are shown in Figs. 10 and 11.

3) Incremental Embedding With the Newly Additive classes: For each individual, 8 images, 11 images, and 10 images in CMU PIE, AR, and EYale B databases, respectively, are used for training, and the remaining images are for testing set.