Learning Semantic Patterns with Discriminant Localized Binary Projections

Shuicheng Yan[1,3], Xiaoou Tang[1,2], and Tianqiang Yuan[1]
[1]Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong
[2]Microsoft Research Asia, Beijing, China
[3]Beckman Institute, University of Illinois at Urbana-Champaign, USA
{scyan,xtang,tqyuan}@ie.cuhk.edu.hk

Abstract

In this paper, we present a novel approach to learning semantic localized patterns with binary projections in a supervised manner. The pursuit of these binary projections is reformulated into a problem of feature clustering, which optimizes the separability of different classes by taking the members within each cluster as the nonzero entries of a projection vector. An efficient greedy procedure is proposed to incrementally combine the sub-clusters by ensuring the cardinality constraints of the projections and the increase of the objective function. Compared with other algorithms for sparse representations, our proposed algorithm, referred to as Discriminant Localized Binary Projections (*dlb*), has the following characteristics: 1) *dlb* is supervised, hence is much more effective than other unsupervised sparse algorithms like Non-negative Matrix Factorization (NMF) in terms of classification power; 2) similar to NMF, *dlb* can derive spatially localized sparse bases; furthermore, the sparsity of *dlb* is controllable, and an interesting result is that the bases have explicit semantics in human perception, like eyes and mouth; and 3) classification with *dlb* is extremely efficient, and only addition operations are required for dimensionality reduction. Extensive experimental results show significant improvements of *dlb* in sparsity and face recognition accuracy in comparison to the state-of-the-art algorithms for dimensionality reduction and sparse representations.

1. Introduction

Subspace learning methods [2][4][20][7] have been widely used in pattern classification owning to their computational simplicity and analytical attractiveness [3]. Most of them, such as Principal Component Analysis (PCA) [8][18], Linear Discriminant Analysis (LDA) [1][13] and the recently proposed Marginal Fisher Analysis (MFA) [22], are holistic, that is, all entries of a projection vector may be nonzero and the computation of each low dimensional feature needs to explore all the features in the original feature space. Methods for sparse representation have been studied to find projection vectors with few nonzero elements. Non-negative Matrix Factorization [10] is the pioneering work towards such a property. It imposes non-negativity constraints in learning the projection vectors. The elements of the projection vectors, *i.e.*, bases, together with the low dimensional representations, are all non-negative. This ensures that the basic projection vectors shall be combined to form an image in a non-subtractive way.

NMF has been extended to Non-negative Tensor Factorization (NTF) [5] for handling the data encoded as general tensors; and Wang *et al* proposed the Fisher NMF [21] by adding a term of scatter difference to the objective function of NMF. Recently, Tao *et al.* [17] proposed an algorithm to employ rectangle features for image reconstruction, which substantially enhances the computational efficiency by taking advantage of the *Integral Image*. There were some attempts to utilize bases with ± 1 values for image transformation, such as the projection with Walsh-Hadamard kernel [6]. Also there were some works to develop bases with ± 1 and zero values for image analysis, such Harr wavelets (normalized bases) for image compression, which have ever been extended to Harr-like features and work as weak classifiers for real time face detection [19]. However, most previous dimensionality reduction algorithms on sparse or binary representations are motivated by image reconstruction in an unsupervised manner, hence they are not necessarily optimal in terms of classification power. Up to now, it is still not clear how to automatically extract spatially localized features effectively in supervised learning.

To deal with the above problem, we present a novel approach which learns a set of localized binary projections with a user-defined cardinality (defined as the number of non-zero elements) for supervised dimensionality reduction. More specifically, the problem is first formulated as a task to learn orthogonal binary projections with limited cardinality, which is done by optimizing a separability criterion that maximizes the average distance between samples

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

of different classes while minimizing that distance between samples of the same class. Then such a problem is transformed to a problem of feature clustering where the members within each cluster correspond to the nonzero entries of a projection vector. Finally, a greedy procedure is proposed to optimize the above separability criterion by progressively combining the sub-cluster pairs into a single cluster.

The advantages of this new algorithm, called discriminant localized binary projections (*dlb*), stem from the following characteristics : 1) *dlb* is a supervised learning method and the derived sparse bases encode information that has the most discriminating power; 2) the bases are spatially localized and have explicit perceptual semantics associated with the human cognition of an object; 3) the features are extracted by summing local patches, which make them more robust; and 4) the computational complexity is greatly reduced, only addition operations are used in *dlb*.

The remainder of this paper is organized as follows: In section 2, we present the algorithm of *dlb* and discuss its complexity. Experiments on sparse analysis, face recognition and robust analysis are shown in section 3. We conclude with section 4.

2. Discriminant Localized Binary Projections

Assume the sample points are given as $\{x_i | x_i \in \mathbb{R}^m\}_{i=1}^N$ where the corresponding class labels are $\{c_i | c_i \in \{1, ..., N_c\}\}_{i=1}^N$. Denote the sample number of the c-th class as n_c. Since in practice the dimension m is often very high, it is usually necessary to transform the data from the input space to a low-dimensional space to alleviate the problem of the curse of dimensionality. For example, if each data point represents an image, then m shall be equal to the number of image pixels, which is over $10,000$ for an image of moderate size. Many dimensionality reduction techniques have been extensively studied and they have achieved much success in real applications [15].

2.1. Motivations and Problem Statement

Classical dimensionality reduction methods usually find a projection matrix $P = [p_1, p_2, ..., p_d] \in \mathbb{R}^{m \times d}$, which maps the original high dimensional feature $x \in \mathbb{R}^m$ to a low dimensional one $y \in \mathbb{R}^d$ by $y = P^T x$. Generally, there is no constraint imposed on the entries of the projection vectors p_i and all entries in p_i can be nonzero, hence the vectors are holistic. However, psychological and physiological evidence have shown the component-based representations in the brain [10]. The problem of sparse representations has been studied, and the Non-negative Matrix Factorization algorithm [10] is proposed for such a purpose. Though non-negative bases and coefficients can be derived in NMF, similar to the Harr wavelet bases and Walsh-Hadamard kernels, NMF and its variants [5][21] are originally intended for image reconstruction. Hence, they are not necessarily effective for classification tasks.

There exists evidence that humans recognize a face often in a local patch-based manner, such as that a face has a high nose, a small mouth, or big eyes. This encourages us to utilize sparse localized bases for feature extraction. Another observation is that, for general projection vectors with both positive and negative values, the extracted features are easily affected by image misalignment, say, translations or scale variations, and one possible way out of this problem is to require the entries of the projection vectors to be binary, which means that the extracted feature must be the sum of some features.

Motivated by the above analysis, we study the supervised dimensionality reduction problem by imposing the following constraints: 1) the bases, *i.e.* the projection vectors, only consist of binary entries; 2) the transformed features from the projection vectors must be effective in classification, unlike the criterion in NMF which emphasizes reconstruction ability; and 3) the bases are spatially localized and orthogonal to each other. Formally speaking, this problem can be defined as follows:

Problem Definition: Given the sample set $X = \{x_i\}_{i=1}^N$ and the corresponding class labels $\{c_i\}_{i=1}^N$, we search for a set of projection vectors $P = [p_1, p_2, ..., p_d]$ that satisfy

$$
\begin{aligned}
&P = \arg\max_P F(P), \quad s.t. \\
&1.\ p_i(k) = 1 \text{ or } 0,\ i = 1, 2, ..., d; k = 1, 2, ...m \\
&2.\ p_i \perp p_j,\ \forall\, i \neq j \\
&3.\ Card(p_i) \leq N_s,\ \forall\, i
\end{aligned}
$$

where $F(P)$ is the objective function which characterizes the classification performance of the projection matrix P; $Card(p_i)$ is the cardinality (number of non-zero entries) of the projection vector p_i, and N_s is an upper bound of the cardinality.

The objective function $F(P)$ can be defined differently for different purposes, such as the Fisher criterion [1] and the Maximum Margin criterion [11]. In our paper, we present a new objective function motivated by the Nearest Neighborhood method. In this objective function, the separability of each sample is evaluated using the difference between its average distance to the other samples in the same class and that to the samples in other classes, and the final objective function is the sum of these difference values from all samples, that is,

$$
\begin{aligned}
F(P) = \sum_{i=1}^{N} (&- \sum_{c_j = c_i} ||P^T x_i - P^T x_j||^2 / (n_{c_i} - 1) \\
&+ \sum_{c_j \neq c_i} ||P^T x_i - P^T x_j||^2 / (N - n_{c_i})).
\end{aligned} \tag{1}
$$

The above supervised learning problem along with the

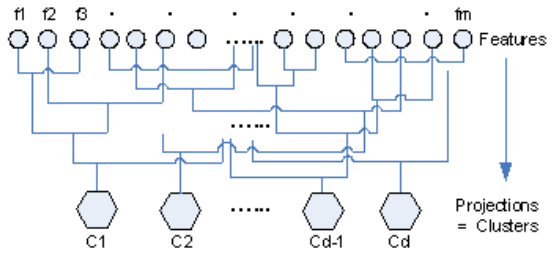


Figure 1. Illustration of the framework for Discriminant Localized Binary Projections in the point of view of clustering.

objective function defined in Eq.(1) is a classical *Integer Optimization* problem. Yet, the number of parameters to be optimized is too large, and it is prohibitive to directly optimize the objective function with traditional techniques for the integer optimization problem; hence it is more desirable to have an efficient procedure which approximates the optimal solution.

2.2. Solution Pursuit

In this subsection, we present a procedure to obtain an approximate solution for the above problem. Since the projection vectors are orthogonal and consist of binary entries, there shall be at most one entry being non-zero within each row of the projection matrix P. Therefore, the above problem can be equally reformulated into a feature clustering problem. Each projection vector p_i of the matrix P corresponds to a feature cluster C_i, and the projection vector p_i acts as an indicator vector for clustering. If the entry $p_i(k)$ is one, then the feature f_k is assigned to the cluster C_i; otherwise not. The corresponding feature clustering problem can be formally formulated as follows:

Problem Reformulation: For a given set of features $\{f_k\}_{k=1}^m$, search for a method that separates them into (d+1) clusters $\{C_0, C_1, ..., C_d\}$ by maximizing the objective function $F(P)$ along with the constraints that $p_i(k)$=1 $\Leftrightarrow$ $f_k \in C_i$, $p_i(k)$=0 $\Leftrightarrow$ $f_k \notin C_i$, $\forall\ i, k$ and the feature number within each cluster is not more than N_s.

In the above clustering process, features in cluster C_0 do not appear in the projection vectors, and they contribute less to the separation of different classes. A local optimum of this problem can be obtained using the *reassignment* method, which has ever been used for the clustering algorithm *K-means* [4]; yet this method is also not practical since we have to recompute the objective function for each reassignment and know the cluster number that is usually unknown in advance. In the following paragraph, we develop a greedy method to progressively combine sub-clusters into a single one, and increase the objective function value step by step while satisfying all constraints.

Greedy Solution. The objective function in Eq. (1) can be rewritten in a trace form as

$$F(P) = Tr(P^T S P) = \sum_{k=1}^{d} p_k^T S p_k \qquad (2)$$

where

$$S = \sum_{i=1}^{N} \left(\frac{1}{N - n_{c_i}} \sum_{c_j \neq c_i} x_{ij} x_{ij}^T - \frac{1}{n_{c_i} - 1} \sum_{c_j = c_i} x_{ij} x_{ij}^T \right) \qquad (3)$$

and $x_{ij} = x_i - x_j$.

Instead of directly optimizing the objective function, we present a greedy solution and assume that the approximate solution is obtained by combining two sub-clusters into a single one progressively meanwhile ensuring that the increase of the objective function value is maximal in each step. In this process, the first two constraints are naturally satisfied, and the last one can be easily met by ensuring that the feature number of the combined cluster is less than N_s.

Assume that the projection matrix is initialized as $P^0 = I_m \in \mathbb{R}^{m \times m}$, where I_m is the identity matrix of size m, that is, each feature constitutes a single sub-cluster. In the (t+1)-th step, two sub-clusters are combined into one, i.e., two column vectors of the projection matrix $P^t = [p_1^t, p_2^t, ..., p_m^t]$ are summed into one $p_i^{t+1} \Leftarrow p_i^t + p_j^t$, and

---

1. Initialize each feature f_k as a cluster C_k, namely $P^0 = I_m$, and set the ineffective feature cluster as $C_0 = \emptyset$;
2. Compute the matrix S as in Eq. (3);
3. for $t = 1, 2, ..., m$-d,
   - Select the maximal non-diagonal entry (i, j) of matrix $S^{t-1} (= {P^{t-1}}^T S P^{t-1})$ which satisfies the condition that the number of the nonzero entries within the combined cluster is no more than N_s.
   - if $(p_i^{t-1} + p_j^{t-1})^T S (p_i^{t-1} + p_j^{t-1}) \leq 0$, then put these two features into cluster C_0, namely set $p_i^t = p_j^t = 0$; else if the entry (i, j) of S^{t-1} is not larger than zero, break; else, set $p_i^t = p_i^{t-1} + p_j^{t-1}, p_j^{t-1} = 0$, namely combine the sub-clusters C_i and C_j into cluster C_i and set the cluster $C_j = \emptyset$;
   - $S^t = {P^t}^T S P^t$
4. Reorder the projection vectors according to $p_i^{tT} S p_i^t$ from large to small, and output $P = [p_1, p_2, ..., p_d]$.

Figure 2. Procedure to learn discriminant localized binary projections.

p_j^t is reset as $p_j^{t+1} = 0$. Thus, the objective function value is increased by

$$(p_i^t + p_j^t)^T S(p_i^t + p_j^t) - {p_i^t}^T S p_i^t - {p_j^t}^T S p_j^t \quad (4)$$

$$= 2{p_i^t}^T S p_j^t = 2e_i^T({P^t}^T S P^t)e_j \quad (5)$$

where e_i is an m dimensional binary vector with only one entry having the value of one at the i-th entry.

The above analysis shows that we need only to find the maximal non-diagonal entry of the matrix $S^t = {P^t}^T S P^t$ under the constraint that the feature number within the combined cluster is not larger than N_s. The computation of the maximal non-diagonal entry is efficient; furthermore, the matrix $S^t = {P^t}^T S P^t$ can be incrementally computed as

$$S^{t+1} = {P^{t+1}}^T S P^{t+1} \quad (6)$$

$$= (P^t(I + E_{ji} - E_{jj}))^T S P^t (I + E_{ji} - E_{jj}) \quad (7)$$

$$= (I + E_{ji} - E_{jj})^T S^t (I + E_{ji} - E_{jj}) \quad (8)$$

where E_{ji} is an $m \times m$ binary matrix with only one entry being one at (j, i). Thus it can be efficiently computed with one row and one column operation for matrix S^t.

In each step, if the combined feature derives a negative value at the (i, i) entry of matrix $S^{t+1} = {P^{t+1}}^T S P^{t+1}$, the elements within this cluster are transferred to cluster C_0 and this cluster is emptied. If the calculated non-diagonal entry (i, j) of matrix S^t is negative or zero, the algorithm is terminated and the final clustering result is output. The detailed procedure is listed in Fig. 2, and Fig. 1 illustrates our framework of combining the features into a set of clusters.

2.3. Algorithmic Analysis

In this section, we discuss the complexity of our algorithm, referred to as Discriminant Localized Binary projections (*dlb*) in both learning and classification stages, and show that *dlb* is very efficient in terms of computation. We shall also analyze its advantages compared with other algorithms for dimensionality reduction, especially for sparse representations.

Complexity of the learning stage. The computational cost of *dlb* in the learning stage is divided into two main parts: one for the computation of the matrix S, which has a complexity of $O(N^2m^2) \times T_\times$ where $T_\times$ is the time for a multiplication operation; another for the computation of the largest entry of the non-diagonal matrix ${P^t}^T S P^t$ in all the m-d steps, which has a complexity of $O(m^3) \times T_+$ where T_+ is the time for an addition operation. Therefore, the total complexity of *dlb* is $O(N^2m^2)T_\times + O(m^3)T_+$, which is even less than that of the PCA algorithm, which is $O(Nm^2)T_\times + O(m^3)T_\times$, when $N \ll m$. Moreover, the average distances in Eqn. (1) can be restricted within the k nearest neighbors of each sample and the computation of the S matrix is reduced to $O(Nkm^2) \times T_\times$, hence the *dlb* is very efficient in the learning stage. In all our experiments, we set k to be $n_{c_i} - 1$ in this paper.

Complexity of the classification stage. In obtaining the low dimensional representation of new data for classification, since all the projection vectors p_i are binary and orthogonal to each other, then each feature in the original feature space will be used at most once and only addition operations will be used. Therefore, the corresponding complexity is $m \times T_+$, which is much lower than that of PCA and other holistic algorithms, where the complexity is $md \times T_\times$. In NMF, the multiplication operations are also required, hence NMF is less efficient than *dlb*.

Advantages of *dlb*. Generally, *dlb* comprises the advantages of both supervised algorithms, like LDA, and sparse representation algorithms, like NMF. It is effective for classification and consistent with the way that the human brain understands the world. Furthermore, the binary property of *dlb* brings additional merits: 1) *dlb* is robust to image misalignment owing to the fact that the low dimensional features are acquired by summing a subset of features, instead of linearly combining all features with different weights; and 2) the binary property makes the computation of low-dimensional features faster.

3. Experiments

In this section, we present three sets of experiments to evaluate the effectiveness of our algorithm and compare it with holistic algorithms like PCA and LDA, as well as the sparse representation algorithm NMF. The first set of experiments is designed to compare the sparsity properties of the bases between NMF and *dlb*; the second set of experiments is used to compare the classification performance of *dlb*, PCA [18], NMF [10], and PCA+LDA (U-Subspace) which is similar to the unified subspace [20] method and explores all possible combinations of PCA dimensions and LDA dimensions.

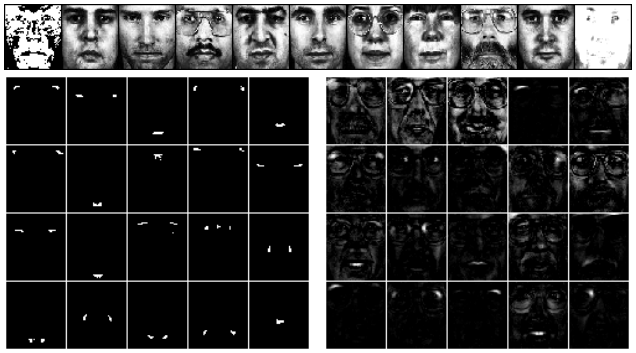
3.1. Sparse Representation

NMF explores sparse non-negative representations for data objects. Its basic objective is to find a set of bases for optimal image reconstruction, hence the bases will usually focus on locations with high gray level values. However, the most effective parts for classification do not always lie in such areas. We compare the results of NMF and *dlb* on three databases, XM2VTS [12] face database, MNIST digital number databases [9], and the CMU PIE [16] database.

The XM2VTS database contains 295 persons and each person has four frontal images each taken in a different session. All the images are aligned by fixing the locations of two eyes and normalizing the size to 72*64 pixels.

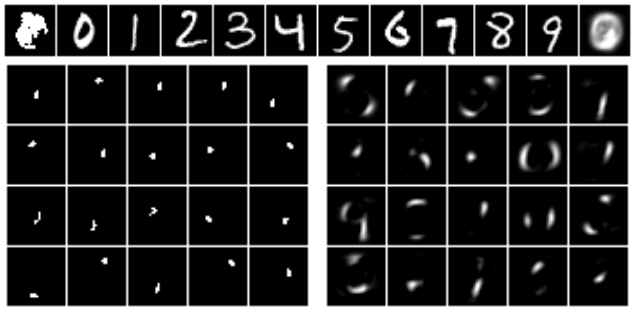
The MNIST digital number database consists of images of handwritten numbers ('0'-'9') extracted from a collection of Dutch utility maps; and we use the 5,000 images in the test set with image size of 28*28 pixels. The values of these images are binary. The CMU PIE (Pose, Illumination, and Expression) database contains more than 40,000 facial images of 68 persons. The images were acquired over different poses, under variable illumination conditions and with different facial expressions. We select nine images for each person as shown in Fig. 5, and there are strong pose variations within the selected images. All the images are aligned by fixing the locations of two eyes and normalizing the size to 64*64 pixels, and 63 persons are used in our experiments due to the data incompleteness of the other five persons.

The learned bases of *dlb* and NMF from the above three databases are displayed in Fig. 3, 4 and 5. The bases of *dlb* are reordered according to their objective function values $p_i^T S p_i$; and if the bases number is larger than 20, the bases of NMF are reordered according to the sums of the projections to the bases from all the samples. We also plot the importance map for *dlb* and NMF, which is defined as the sum of the leading projection vectors and characterizes the importance of the feature for a specific algorithm. All the projection vectors are linearly scaled to within $[0, 255]$ for display. From these results, we have the following observations: 1) when the faces are well aligned, like in the XM2VTS database, NMF can derive localized features; but there still exist holistic bases as shown in Fig. 3; 2) for binary images, like the MNIST database, the bases of NMF are sparse; 3) for face images with different poses, like in the CMU PIE database, the bases of NMF are not satisfactory in sparsity, and many distorted faces are derived; 4) in all of the three cases, *dlb* consistently outputs sparse and localized bases, and the bases from face images may have explicit perception semantics, like eyes and mouths; and 4) the importance maps of NMF and *dlb* are different, or even contrary to each other, which indicates that the most representative features are usually not the best in classification capacity and also show why *dlb* can work better than NMF in classification tasks. Also, we plot the bases with different cardinalities in Fig. (6) with the 210 images of 70 persons from the FERET [14] database. From these results, we can observe that when the cardinality of the projection vector is small, the first six bases mainly focus on the eye area, and when the cardinality increases, the mouth and nose are also included. When the cardinality is small enough, the bases are localized and have explicit perceptive semantics; and when the cardinality increases, the bases exhibit high level structure information of a face, and multiple components are combined to constitute a basis.



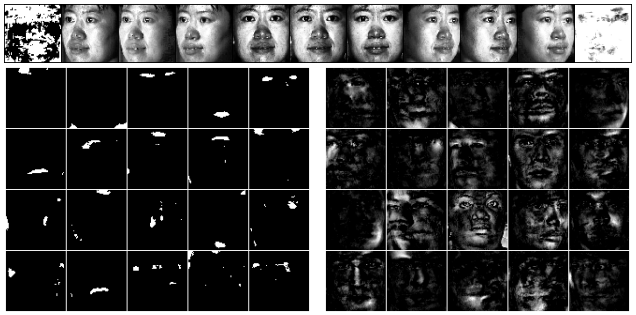
(a) *dlb* (b) NMF

Figure 3. The first 20 bases of *dlb* (bottom left) and NMF (bottom right) from the XM2VTS database with well aligned frontal faces. The top row images are: importance map image of *dlb*, nine sample images from the XM2VTS database, and the importance map image of NMF.



(a) *dlb* (b) NMF

Figure 4. The first 20 bases of *dlb* (bottom left) and NMF (bottom right) from the MNIST digital number database. The top row images are: importance map image of *dlb*, ten sample images from the MNIST database, and the importance map image of NMF.



(a) *dlb* (b) NMF

Figure 5. The first 20 bases of *dlb* (bottom left) and NMF (bottom right) from the PIE database with multi-view faces. The top row images are: importance map image of *dlb*, nine sample images from the PIE database, and the importance map image of NMF.

3.2. Face Recognition

We evaluate the classification power of the derived low-dimensional representations from PCA, U-Subspace, NMF,

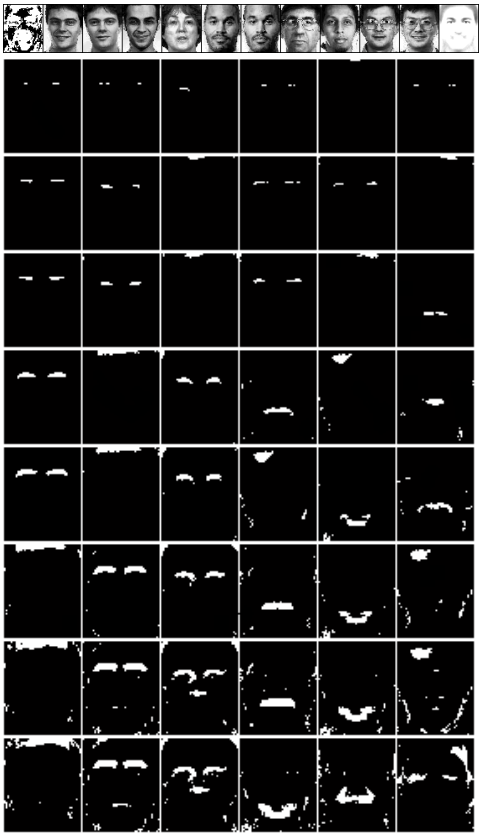
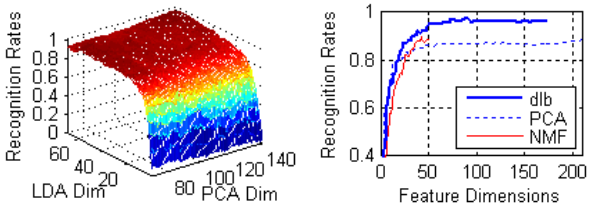


Figure 6. The first six bases (from left to right) of *dlb* from the FERET database with different cardinalities (from the second row to the bottom row with the cardinality from 4 to 180); and the top row are the importance map of *dlb*, ten sample images and the importance map of NMF. The results show that the bases have explicit human perceptive semantics, like eyes, mouth and nose; and when the cardinality increases, the bases can show high level structure information of a face by combining multiple components.

and *dlb* on two face databases: FERET [14] and CMU PIE. In this experiment, we used 70 persons from the FERET database with six images for each person and the images are aligned by fixing the locations of two eyes to the size of 56×46 pixels. Three images per person are randomly selected for training and the remaining three images are for testing. In CMU PIE database, the nine images for each person as shown in Fig. 5 are used for training and other twelve images each person are selected from poses 05, 07, 09 and 29 with three different illuminations for testing. The detailed experimental results are displayed in Fig. 7 and Fig. 8. In all the experiments, the projection vectors are learned from the training images and the Nearest Neighbor approach is used for classification; we explored all the possible feature dimensions for all the algorithms and report the best results. For a fair comparison, we implemented the Unified Subspace (U-Subspace) method, similar to the unified subspace method in [20], by exploring all possible

PCA	U-Subspace	NMF	*dlb*
88.6%(206)	96.2%(136,66)	89.5%(42)	97.1%(73)
$\approx$ 2.85 ms	$\approx$ 0.91 ms	$\approx$ 0.58 ms	$\approx$ 0.01ms



(a) U-Subspace (b) PCA, NMF, *dlb*

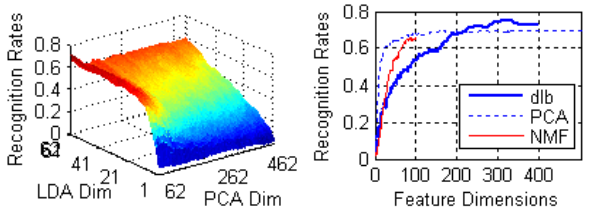
Figure 7. Face recognition results on the FERET database. Note that, in the table above the two images, the first line lists the best results from the four algorithms, and second line lists the approximate time (in Matlab 7.0 on PIV computer with CPU 3.0G) required for computing the optimal low dimensional representation with corresponding feature dimensions in the brackets (for U-Subspace method, the two numbers in the bracket are PCA dimension and LDA dimension respectively); the left image shows the recognition rates of U-Subspace algorithm on different combinations of PCA and LDA dimensions; and the right image shows the face recognition rates of PCA, NMF and *dlb* on different feature dimensions.

combinations of PCA dimensions and LDA dimensions; for NMF, we implemented the versions with bases numbers of 20, 50, 100, 200 and reported the best results; for *dlb*, we implemented the versions with cardinality N_s as 5, 15 and 25 respectively and reported the best results. From these results, we have the following conclusions: 1) U-Subspace consistently outperforms PCA and NMF when all feature dimensions are explored; 2) *dlb* is comparable to the U-Subspace algorithm in recognition accuracy with significantly improved processing speed; and 3) when there are large pose variations, NMF is not so good as PCA, yet it can outperform PCA in the cases with frontal faces. We also evaluate the performance of *dlb* with different cardinalities in comparison to that of NMF with different numbers of bases for training, and the results on the FERET database are shown in Fig. 9. These results show that *dlb* is relatively more robust to model parameter variations than NMF.

4. Discussions and Future Work

In this paper, we proposed a novel supervised dimensionality reduction algorithm which pursues discriminant localized binary projections. The pursuit of these projections is simplified into a supervised feature clustering problem; and a greedy procedure is proposed to hierarchically combine projection vector pairs into single ones. *dlb* can derive localized bases with explicit perceptual semantics and strong classification power. The binary property of *dlb* brings the advantages of low computational complexity and robustness. One possible future work in this direction is to train

PCA	U-Subspace	NMF	*dlb*
69.7%(264)	72.0%(62,61)	65.6%(88)	75.4%(314)
≈ 5.75 ms	≈ 1.34 ms	≈ 1.93 ms	≈ 0.04ms



(a) U-Subspace (b) PCA, NMF, *dlb*

Figure 8. Face recognition results on CMU PIE database. The contents shown in the table and two images are similar to those at Fig. 7

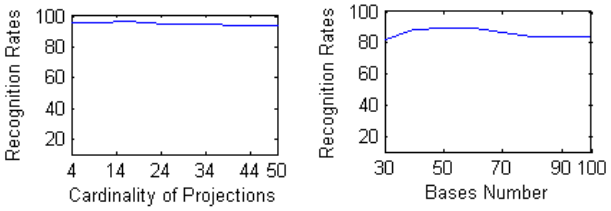


Figure 9. Face recognition rates of *dlb* with different cardinality numbers, and those of NMF with different basis numbers, for training on the FERET database.

a specific classifier within each local patch instead of manually separating the image into multiple patches for parts-based face recognition. Another possible extension is to explore other forms of objective functions in *dlb* for better classification power while preserving localized and binary properties.

Acknowledgement

The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region. The work was conducted at the Chinese University of Hong Kong.

References

[1] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 2002. 1, 2

[2] F. Chung. Spectral graph theory. *Regional Conferences Series in Mathematics*, 92, 1997. 1

[3] J. M. *et al*. Comparison of face verification results on the xm2vts database. *Proceddings of International Conference on Pattern Recognitions*, 4:858–863, 2000. 1

[4] K. Fukunaga. Introduction to statistical pattern recognition. *Academic Press, second edition*, 1991. 1, 3

[5] T. Hazan, S. Polak, and A. Shashua. Sparse image coding using a 3d non-negative tensor factorization. *International Conference on Computer Vision (ICCV) Beijing, China*, 2005. 1, 2

[6] Y. Hel-Or and H. Hel-Or. Real-time pattern matching using projection kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):1430–1445, 2005. 1

[7] A. Hyvärinen, J. Karhunen, and E. Oja. Independent component analysis. *John Wiley & Sons*, 2001. 1

[8] I. Joliffe. Principal component analysis. *Springer-Verlag, New York*, 1986. 1

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 4

[10] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999. 1, 2, 4

[11] H. Li, T. Jiang, and K. Zhang. Efficient and robust feature extraction by maximum margin criterion. *Advances in Neural Information Processing Systems 16*, 2004. 2

[12] J. Luettin and G. Maitre. Evaluation protocol for the extended m2vts database (xm2vts). *DMI for Perceptual Artificial Intelligence*, 1998. 4

[13] Martinez and A. Kak. Pca versus lda. *IEEE Transactions on Pattern Analysis and Ma-chine Intelligence*, 23(2):228–233, 2001. 1

[14] I. Philips, H. Wechsler, J. Huang, and P. Rauss. The feret database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16:295–306, 1998. 5, 6

[15] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *Proceddings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005. 2

[16] T. Sim and S. Baker. The cmu pose, illumination, and expression database. 4

[17] H. Tao, R. Crabb, and F. Tang. Non-orthogonal binary subspace and its applications in computer vision. *Proceedings of Computer Vision and Pattern Recognition Conference*, 2005. 1

[18] M. Turk and A. Pentland. Face recognition using eigenfaces. *IEEE Conference on Com-puter Vision and Pattern Recognition, Maui, Hawaii*, 1991. 1, 4

[19] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. 1

[20] X. Wang and X. Tang. A unified framework for subspace face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1222–1228, 2004. 1, 4, 6

[21] Y. Wang, Y. Jiar, C. Hu, and M. Turk. Fisher non-negative matrix factorization for learning local features. *Asian Conference on Computer Vision, Korea, January 27-30*, 2004. 1, 2

[22] S. Yan, D. Xu, B. Zhang, and H. Zhang. Graph embedding: A general framework for dimensionality reduction. *Proceedings of Computer Vision and Pattern Recognition Conference*, 2005. 1