# Linear Projection Methods in Face Recognition under Unconstrained Illuminations: A Comparative Study

Qi Li† Jieping Ye‡ Chandra Kambhamettu†
† Video/Image Modeling and Synthesis Lab
Computer Information & Science, University of Delaware, Newark, DE, 19716
{qili,chandra}@cis.udel.edu
‡ Computer Science & Engineering, University of Minnesota, Minneapolis, MN, 55455
jieping@cs.umn.edu

## Abstract

*Face recognition under unconstrained illuminations (FR/I) received extensive study because of the existence of illumination subspace. [2] presented a study on the comparison between Principal component analysis (PCA) and subspace Linear Discriminant Analysis (LDA) for this problem. PCA and subspace LDA are two well-known linear projection methods that can be characterized as trace optimization on scatter matrices. Generally, a linear projection method can be derived by applying a specific matrix analysis technique on specific scatter matrices under some optimization criterion. Several novel linear projection methods were proposed recently using Generalized Singular Value Decomposition or QR Decomposition matrix analysis techniques [10, 17, 11]. In this paper, we present a comparative study on these linear projection methods in FR/I. We further involve multiresolution analysis in the study. Our comparative study is expected to give a relatively comprehensive view on the performance of linear projection methods in FR/I problems.*

## 1 Introduction

There have been extensive works on face recognition under unconstrained illuminations because of the existence of illumination subspace which has been shown by the empirical study [7, 5, 2] and the theoretical study [1, 12]. Among different choices of illumination subspaces, 3-D, 4-D, 5-D, 7-D, and 9-D have received extensive attention [7, 5, 2, 1]. [2] studied 3-D illumination cones where face surface is assumed to be Lambertian without self-shadowing. Most recent works focused on 9-D illumination subspaces. More specifically, [6] presented a sophisticated model-based method; [8] presented 9 pointlights by hardware configuration; [18] presented a statistical method (bootstrap face models are used to capture the statistics of illuminations). The limitation of these methods is the involvement of 3-D face models or special physical configuration, which makes it difficult to apply to more general face recognition problems. For example, under unconstrained expressions and illuminations, 3-D face models involved have to contain different facial expressions.

[2] presented a study on the comparison between Principal component analysis (PCA) [15] and subspace Linear Discriminant Analysis (LDA) [14] in face recognition under unconstrained illuminations. PCA and subspace LDA are two well-known linear projection methods that can be characterized as trace optimizations on scatter matrices. Generally, a linear projection method can be derived by applying a specific matrix technique on specific scatter matrices under some optimization criterion. The projection criterion used in PCA is to choose the directions along which training data has maximum variation to construct the reduced subspace. The basic criterion of LDA is to maximize the between-class scatter and minimize within-class scatter. Classical LDA has singularity problem (i.e., it will fail if both scatter matrices are singular). Subspace LDA, also called PCA+LDA, is a widely used implementation of LDA. Subspace LDA uses PCA to intermediately reduce the dimension of original data so that the singularity problem can be overcome. But the intermediate dimension reduction can cause information loss.

Recently, several novel linear projection methods were proposed. One is called Orthogonal Centroid Method (OCM) that aims to maximize the between-class distance (by orthogonizing the centroids of classes/subjects) [11]. This method is simple and efficient. LDA/GSVD and LDA/QR are two novel implementations of LDA [10, 16]. By applying Generalized Singular Value Decomposition (GSVD), the former method can overcome the singularity

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)
1063-6919/04 $20.00 © 2004 IEEE

problem without intermediate dimension reduction. By applying QR Decomposition, the latter method converts an eigen decomposition on original scatter matrices to the one on "reduced" scatter matrices, and thus distinctly reduces the cost in computing projection matrices. LDA/QR, as well as PCA+LDA, is a two-stage LDA.

This paper presents a comparative study on different linear projection methods in face recognition under unconstrained illuminations (FR/I). It is a development of the work in [2], where a comparison between PCA and PCA+LDA in FR/I was presented. It is worthy to note that FR/I is a very good testbed in comparing the effectiveness of linear projection methods because of the existence of illumination subspace. We are concerned with the time complexity in this study since an FR/I technique with low time/space complexity and no involvement of 3-D models or physical configuration is expected to target general face recognition problems more easily.

We further involve multiresolution analysis (MRA) [9, 4]. [3] applied PCA+LDA to waveletfaces that refer to faces in the low-frequency domain. [3] did not address FR/I problems.

The experiments in our study were done on CMU-PIE and YaleB [13, 6]. Several novel conclusions are led by this comparative study, including:

- LDA/GSVD, the discriminant method without inversion on scatter matrices, always achieves highest accuracy in different testing scenarios, and it is insensitive to the use of MRA.

- LDA/QR, the discriminant method using QR Decomposition, obtains significantly improved performance with the use of MRA. (The integration of LDA/QR and MRA approximates the accuracies achieved by LDA/GSVD.)

- OCM, the method with the only effort in maximizing between-class distance, outperforms PCA (with or without integration of MRA).

The rest of the paper is organized as follows. Section 2 reviews linear projection methods. Section 3 presents the basics of MRA. Section 4 presents the experimental results and analysis. We further have a discussion on the use of interpolation scheme in constructing training sets for FR/I and the use of longer wavelets in Section 5. Conclusions are given in Section 6.

# 2 Linear projection methods

In this section, we will review Orthogonal Centroid Method and different implementations of LDA. (Since PCA is well-known and technically simple, it is not reviewed here.) These linear projection methods can be unified as a general trace optimization problem. The different use of scatter matrices and different matrix analysis techniques distinguish the functionality and power of these linear projection methods.

In the following illustration, we denote $N$ to be the number of training instances, $n$ to be the dimension of an instance item, and $k$ to be the number of subjects/individuals. Furthermore, between-class scatter matrix $S_b$ and within-class scatter $S_w$ are defined as follows:

$$\begin{aligned} S_b &= \frac{1}{N}\sum_{i=1}^{k} N_i(m_i - m)(m_i - m)^T = \frac{1}{N}H_b H_b^T, \\ S_w &= \frac{1}{N}\sum_{i=1}^{k}\sum_{x\in A_i}(x - m_i)(x - m_i)^T = \frac{1}{N}H_w H_w^T, \end{aligned} \quad (2.1)$$

where

$$\begin{aligned} H_b &= [\sqrt{N_1}(m_1 - m), \cdots, \sqrt{N_k}(m_k - m)] \in R^{n\times k}, \\ H_w &= [A_1 - m_1 \cdot e_1, \cdots, A_k - m_k \cdot e_k] \in R^{n\times N}, \end{aligned} \quad (2.2)$$

$e_i = (1, \cdots, 1) \in R^{1\times N_i}$, $A_i$ is the data matrix for $i$th class, $m_i$ is the centroid of $i$th class, and $N_i$ is the number of instances in $i$th class.

## 2.1 Orthogonal Centroid Method (OCM)

OCM aims to maximize the separability of different classes [11]. OCM is based on QR Decomposition, and it utilizes the centroid information of training data, both of which lead to the efficiency of OCM in contrast to PCA. (We will find that OCM also achieves higher accuracy in FR/I than PCA, shown in the later experimental part.) More specifically, assume $C \in \mathbb{R}^{n\times k}$ is a matrix, each column of which is the centroid of training instances of a subject. OCM uses the following optimization criterion,

$$G = \arg\max_{G^T G = I} \operatorname{trace}(G^T S_b G). \quad (2.3)$$

The solution to (2.3) can be obtained through QR Decomposition as follows. Let $C = QR$ be the QR Decomposition on $C$, where $Q \in R^{n\times k}$, $R \in \mathbb{R}^{k\times k}$. Then $G = QW$ for any orthogonal matrix $W \in \mathbb{R}^{k\times k}$ solves the optimization in (2.3). The dimension of projected subspace via OCM is $k$, the number of subjects. The time complexity of OCM is $O(nNk)$.

## 2.2 Classical LDA

Classical LDA is commonly found by solving the trace optimization of the following,

$$G = \arg\max_{G} \operatorname{trace}\left((G^T S_w G)^{-1}(G^T S_b G)\right), \quad (2.4)$$

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)
1063-6919/04 $20.00 © 2004 IEEE

The optimization criterion in (2.4) is equivalent to the following generalized eigen problem,

$$S_b x = \lambda S_w x, \text{ for } \lambda \neq 0. \quad (2.5)$$

The solution can be obtained by solving an eigen problem on matrix $S_w^{-1} S_b$. There are at most $k-1$ non-zero eigenvalues, since the rank of the matrix $S_b$ is bounded by $k-1$. Therefore, the reduced dimension by classical LDA is at most $k-1$. A way to solve the eigen problem is to apply SVD on the scatter matrices. Classical LDA will fail if the scatter matrix $S_w$ is singular, which occurs frequently in face recognition where face images (even the cropped ones) usually have extremely high dimensionality.

### 2.3 PCA+LDA

One way to overcome the singularity of $S_w$ is to use PCA to reduce the dimension of the original data before classical LDA is applied. This is known as PCA+LDA or subspace LDA [19]. It is a popular face recognition technique [2, 14]. The choice of the reduced dimension for PCA is the biggest challenge in using PCA+LDA. If the reduced dimension is large, the eigen problem in the discriminant stage will be expensive (note that the time complexity is $O(N^2 n)$), and unstable because of the high dimensionality. If it is too small, it may not provide sufficient discriminant information. Generally, PCA+LDA is an expensive method when the number of subjects or the number of training instances is large.

### 2.4 LDA/GSVD

A recent work on overcoming singularity problem in LDA is the use of Generalized Singular Value Decomposition (GSVD) [10, 16]. This method is named LDA/GSVD; it computes the solution exactly (without losing any information), because the inversion of the matrix $S_w$ can be avoided with the use of GSVD. It was shown to outperform PCA, OCM, and PCA+LDA on text classification problem [10, 16]. The time complexity of the LDA/GSVD is $O((N+k)^2 n)$ [10, 16], thus it is also an expensive method.

### 2.5 LDA/QR

[17] presented a novel LDA implementation, namely LDA/QR. LDA/QR contains two stages. The first stage is to maximize separability between different classes and thus has similar target as OCM. The second stage of LDA/QR incorporates the within-class scatter information by applying a relaxation scheme to $W$. The final optimization problem is exactly the same one as in classical LDA. However, when applied to matrices of much smaller size, LDA/QR runs more efficiently and stably.

Specifically, we would like to find a projection matrix $G$ such that $G = QW$, for any matrix $W \in \mathbb{R}^{k \times k}$, hence $W$ is not required to be orthogonal. The original problem of finding $G$ is now equivalent to computing $W$. In the first stage, LDA/QR chooses $W$ to be any orthogonal matrix, by omitting the within-class scatter. It considers both the between-class and within-class scatters in the second stage as follows. Since

$$\begin{aligned} G^T S_b G &= W^T (Q^T S_b Q) W, \\ G^T S_w G &= W^T (Q^T S_w Q) W, \end{aligned} \quad (2.6)$$

the original problem of finding optimal $G$ is equivalent to finding $W$, such that

$$W = \arg\max_W \text{trace}\left( (W^T \tilde{S}_b W)^{-1} (W^T \tilde{S}_w W) \right), \quad (2.7)$$

where $\tilde{S}_b = Q^T S_b Q$ and $\tilde{S}_w = Q^T S_w Q$. Note that $\tilde{S}_b$ has a much smaller size than the original scatter matrix $S_b$ (similarly for $\tilde{S}_w$), which is the main reason why LDA/QR preserves the same time complexity as the first stage. The optimization in (2.7) now can be solved efficiently by solving a small eigen problem on $\tilde{S}_b^{-1} \tilde{S}_w$. The time complexity of LDA/QR is the same as OCM's, i.e., $O(nNk)$.

### 2.6 Time/space complexity

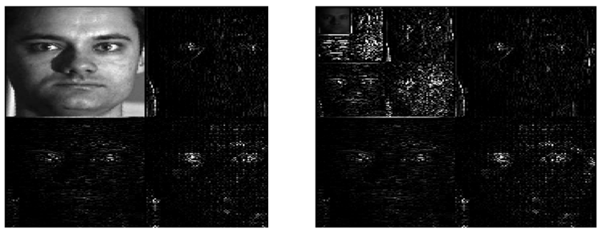
The time/space complexity of the linear projection methods is presented in Table 1. PCA, PCA+LDA and LDA/GSVD are uniformly more expensive than OCM and LDA/QR. Note that MRA is of low time complexity.

| Methods | Time Complexity | Space Complexity |
|---|---|---|
| PCA | $O(N^2 n)$ | $O(nN)$ |
| OCM | $O(nNk)$ | $O(nk)$ |
| PCA+LDA | $O(N^2 n)$ | $O(nN)$ |
| LDA/GSVD | $O((N+k)^2 n)$ | $O(nN)$ |
| LDA/QR | $O(nNk)$ | $O(nk)$ |

**Table 1. Complexity comparison. $N$ is the number of training data points, $n$ is the number of dimensions, and $k$ is the number of classes.**

## 3 Multiresolution analysis (MRA)

We now give a brief review on multiresolution analysis (MRA) (from the perspective of wavelet analysis). More details can be found in [9]. We start from MRA on a 1D signal space and then on a 2D signal (i.e., image) space. Assume $V_0$ is the highest resolution of a digitized 1D signal space. MRA of $V_0$ is the ladder of subspaces of $V_0$:



(a) Two layers (b) Four layers

**Figure 1. An example of the MRA ladder in image plane.**

$$\begin{array}{ccccccc} V_{n_0} & \subset & \cdots & \subset & V_1 & \subset & V_0 \\ \oplus & & \cdots & & \oplus & & \\ W_{n_0} & & \cdots & & W_1 & & \end{array}$$

where $V_{n_0}$ is the lowest-resolution subspace, and $W_j$ is the complement of $V_j$ in super space $V_{j-1}$ (i.e., $V_{j-1} = V_j \oplus W_j$). With MRA, $V_0$ can be expressed as $V_{n_0} \oplus W_{n_0} \cdots \oplus W_1$.

The MRA of a digitized image space can be obtained by introducing the tensor products of 1D signal subspaces. More specifically, assume $\mathbf{V}_{j-1}$ is a resolution subspace of an image space that can be represented as the tensor product of $V_{j-1}$ and $V_{j-1}$. We have

$$\begin{aligned} & \mathbf{V}_{j-1} = V_{j-1} \otimes V_{j-1} = (V_j \oplus W_j) \otimes (V_j \oplus W_j) \\ = \;& (V_j \otimes V_j) \oplus [(W_j \otimes V_j) \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j)] \\ = \;& \mathbf{V}_j \oplus \mathbf{W}_j, \end{aligned}$$

where $\mathbf{W}_j = [(W_j \otimes V_j) \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j)]$ describes the subspace of high-frequency information in resolution $j$. Given an image, these high-frequency subspaces capture the local changes of intensities along horizontal, vertical, and diagonal directions respectively. They are commonly denoted as $LH_j$, $HL_j$ and $HH_j$ bands respectively (here $L$ refers to low-frequency, and $H$ refers to high-frequency). Fig. 1 shows an example of the multiresolution ladder of a given image plane where original images are decomposed into instances in four consecutive subspaces. In the following experimental study, we will see that MRA surprisingly enriches the behaviors of linear projection methods. In using MRA, we always throw away the low-frequency information $LL_{n_0}$.

## 4 Experiments

In this section, we will first present the datasets and associated validation schemes, then the results and observations, and finally the analysis.

The first 100 principal components are used to construct projection subspaces in PCA method, and $k-1$ discriminant components are used in LDA methods. The wavelet used to implement MRA is Haar. For convenience, we denote n-Layer to be the concatenation of high frequency information of n-layers. (1-Layer indicates no MRA used.) Similar to the most previous works on face recognition, nearest neighbor is used as the classifier in the face identification stage.

### 4.1 Datasets and validation schemes

**PIE/Pose27:** A subset of CMU-PIE face dataset [13]. CMU-PIE contains 68 subjects/individuals each of which has 13 different poses and 43 illumination conditions, 4 different facial expressions. The dataset PIE/Pose27 used in this paper contains entire subjects each of which has 21 illumination conditions (indexed from 2nd to 22nd) and frontal pose (indexed by 27th). Fig. 2 shows 21 illumination instances of a subject in PIE/Pose27. These images are normalized.

All instances of PIE/Pose27 are stored in a raw data matrix $M$ based on naming locality. For example the rows from 1 to 21 are instances of the first subject, the rows from 22 to 42 are instances of the second subject, and so on. We choose instances by every $m$ rows in $M$ to form the training set (and use the rest for testing). For example, if $m$ is 11, then the 1st, 12th, 23rd, 34th,... row vectors will be chosen to form the training sets. Note that 1st and 12th rows are associated with the 2nd, 13th illumination instances of the first subject respectively, whereas the 23rd, and 34th rows are associated with the 3rd and 14th illumination instances of the second subject respectively. The illumination instances/conditions of different subjects is not necessarily the same one, which can bring flexibility in handling FR/I problems. We will average identification accuracies of all tests of the same $m$. Thus this validation can be considered as "reversed" $n$-fold validation. (n-fold validation uses $\frac{n-1}{n}$ pieces to train, and the remaining $\frac{1}{n}$ pieces to test. Reversed $n$-fold switches the training set and the test set in $n$-fold validation.)

**YaleB/Pose00:** A subset of Yale B face dataset [6]. Yale B contains 10 subjects, 9 poses, and 64 different illumination conditions in each pose. The 64 illumination conditions are further subdivided into 5 subsets. The extremity of illuminations (measured by the azimuth and elevation of the light sources) is increased along with the index of these subsets (so subset 5 contains the most extreme illumination conditions, and subset 1 contains the least). YaleB/Pose00 collects all instances of the frontal pose (indexed by 00). Similar to the previous works [6, 8, 18], we are also interested in the first four subsets. The training set for YaleB/Pose00 always consists of one instance from subset 1 and seven instances from subset 4. These instances are randomly cho-

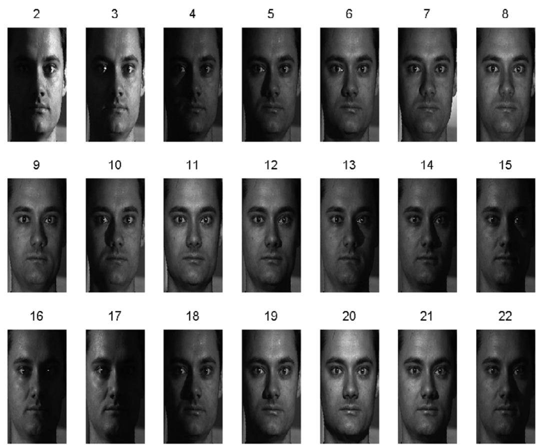



**Figure 2. 21 illumination instances of a subject in PIE/Pose27 (after normalization).**

sen. The final recognition accuracy is the average of different runs. Note that this scheme is essentially consistent with the one used in PIE. We have more discussion on the use of this kind of interpolation schemes in using the linear projection methods in the next section. There are three different test sets: subset 1& 2, subset 3, and subset 4.

### 4.2 Results and observations

Fig. 3 shows the accuracy comparison between using no MRA and MRA to PIE/Pose27. We have the following observations:

- No MRA vs. MRA/Haar. From the view of PCA and OCM, MRA/Haar helps to improve the recognition accuracies, especially in low illumination subspaces. For example, the recogntion accuracies achieved by PCA on 3-D illumination subspace are 60% and 86% using no MRA and MRA/Haar respectively. However, the effectiveness of MRA/Haar differs when it is combined with different LDA implementations. MRA/Haar improves the performance of LDA/QR (say from 90% to 95% on 3-D illumination subspace), while usually degrading PCA+LDA's.

- PCA vs. OCM. The interesting observation is that OCM behaves much better than PCA on low illumination subspaces (say 3-D, 4-D) via MRA/Haar. For example, OCM achieves 95% accuracy rate on 3-D subspace via MRA/Haar, whereas PCA only achieves 87% accuracy rate under the same senario.

- PCA+LDA vs. LDA/QR vs. LDA/GSVD. LDA/GSVD outperforms the other two. Actually, it always performs the best among all linear projection methods considered in this study.

- 3-D vs. 4-D vs. 5-D vs. 7-D vs. 10-D. Without MRA, the dramatic accuracy change occurs at 5-D subspace (i.e., from 5-D to 7-D) from the standpoint of PCA and OCM, whereas there is no distinct accuracy change from the view of LDA methods. With MRA, the accuracy jump occurs at 4-D subspace.

The results on the use of different MRA layers are shown in Fig. 4 from which we have the following observations:

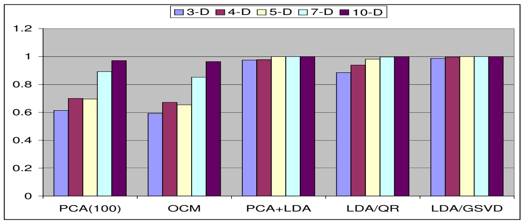
- 1-Layer vs. 2-Layer vs. 3-Layer vs. 4-Layer. With more numbers of MRA layers involved, the recognition accuracies via different methods are increasing. The increasing trends via PCA and OCM are easier to observe than the ones via LDA implementations. For example, the accuracy achieved by PCA in 3-D illumination subspaces increases from 58% to 89% as the number of MRA layers is increasing from 1 to 4. From the opposite view, we equivalently conclude that LDA methods are less sensitive to the number of MRA layers than PCA and OCM.

- PCA vs. OCM. OCM distinctly outperforms PCA, especially on low illumination subspaces, such as 3-D or 5-D. For example, in the case of one layer, PCA achieves accuracy of 57% using 3-D training illumination subspace, while OCM achieves accuracy of 82% using the same training illumination subspace.

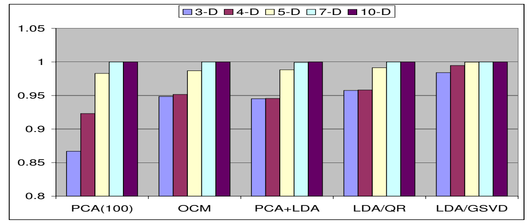
- PCA+LDA vs. LDA/QR vs. LDA/GSVD. LDA/GSVD always achieves highest accuracy. The other two have similar performance. LDA/GSVD is also most stable to the dimensionality of illumination subspaces. For example, in the case of three and four layers, there is a noticeable accuracy change from 4-D subspace to 5-D subspace in all methods except LDA/GSVD.

- 3-D vs. 4-D vs. 5-D vs. 7-D vs. 10-D. 5-D illumination subspace is sufficient to provide good approximation of entire spaces from the perspective of LDA/GSVD (which always achieves 100% accuracy rates in all cases); whereas 7-D illumination subspace is needed to provide 100% accuracy rate from the perspective of PCA+LDA and LDA/QR. There is a distinction between 4-D and 5-D illumination subspace from the perspective of linear projection methods.

Now we present the experimental results on YaleB/Pose00. LDA/GSVD is one of our focus because of the superiority of LDA/GSVD (with or without

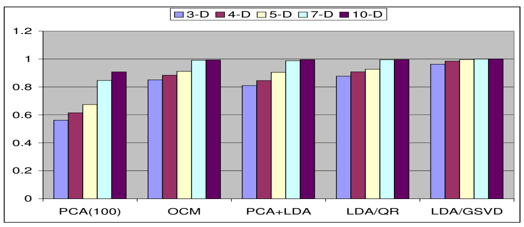



(a1) No MRA

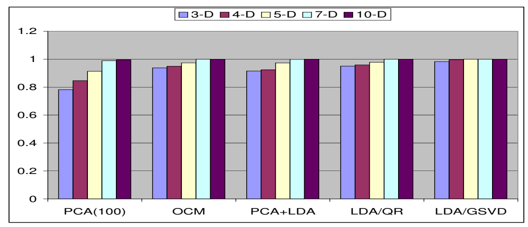



(a2) MRA/Haar

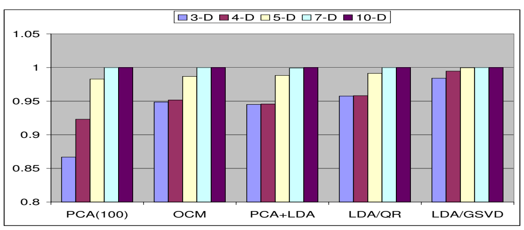
**Figure 3. Comparision between using no MRA and MRA. (The five bars in each projection method are associated with the accuracies using 3-, 4-, 5-, 7-, and 10-D subspaces respectively.)**



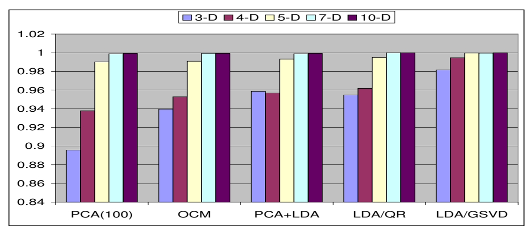

(a1) MRA/1-layer




(a2) MRA/2-layer




(a3) MRA/3-layer




(a4) MRA/4-layer

**Figure 4. Recognition accuracies achieved by different MRA layers of PIE/Pose27.**

MRA) on PIE/Pose27. We will also give attention to LDA/QR because it has much lower time complexity than LDA/GSVD and attains competitive performance. Detailed analysis of the superiority of LDA/GSVD and the tradeoff between LDA/GSVD and LDA/QR will be presented later. For the convenience of evaluating our methods, we also list three best results in the literature on face recognition under unconstrained illuminations [6, 8, 18]. To be consistent with the result presentation in previous work on YaleB, we now use error rate (rather than recognition accuracy) to measure the performance of the recognition methods. Table 2 presents the results that lead to the following observations:

- LDA/GSVD achieves very good performance. LDA/GSVD without MRA is ranked the second best method among all listed methods. Here the use of MRA slightly degrades the performance of LDA/GSVD.
- MRA significantly improves the performance of LDA/QR. Even though LDA/QR with MRA is worse than LDA/GSVD, it is competitive with the Bootstrap and 9PL methods.

We can see that the results on YaleB/Pose00 is consistent with the ones on CMU-PIE/Pose27.

### 4.3 Analysis

**Superiority of LDA/GSVD** An important conclusion in our study on linear projection methods in face recognition under unconstrained illuminations is the superiority of

| Method | Subset 1 & 2 | Subset 3 | Subset 4 |
|---|---|---|---|
| LDA/QR w MRA | 0.5 | 3.3 | 2.1 |
| LDA/QR w/o MRA | 3.2 | 19.1 | 5.7 |
| LDA/GSVD w MRA | 0.5 | 1.7 | 2.1 |
| LDA/GSVD w/o MRA | 0.0 | 0.0 | 1.4 |
| Bootstrap[18] | 0.0 | 0.3 | 3.1 |
| 9PL[8] | 0.0 | 0.0 | 2.8 |
| Cones-cast[6] | 0.0 | 0.0 | 0.0 |

**Table 2. Comparison of recognition methods on YaleB/Pose00. (Measured by error rate in percentage)**

LDA/GSVD over PCA+LDA and LDA/QR. LDA/GSVD not only always achieves the highest accuracies, but also is insensitive to the dimensionality of illumination subspaces and the different number of MRA layers. Thus, LDA/GSVD is much more robust to the different representations of training illumination subspaces than PCA+LDA and LDA/QR.

We believe that two major factors contribute to the superiority and robustness of LDA/GSVD in face recognition under unconstrained illumination problem. The first is the absence of inverse operation on $S_b$ in LDA/GSVD. Without the inverse operation on $S_b$, the bias that may be introduced by different choices of training illumination subspaces and data disturbances can be successfully avoided. The second factor is the existence of illumination subspaces that simplifies the structure of between-class and within-class scatters. Since LDA/GSVD is based on the optimal criterion in interpreting the principle of maximizing between-class distance and minimizing the within-class distance simultaneously [16], whereas PCA+LDA and LDA/QR are approximate solutions [17], we are not surprised to observe its superiority in FR/I problem. It is important to note that the distinct superiority may not always occur if the structure of between-class and within-class scatters is complex.

**Trade-off between LDA/GSVD and LDA/QR** If a face recognition problem only concerns the recognition accuracy (with no concern about the training efficiency, or scalability to the face database), LDA/GSVD should be the number one choice. However, in reality, both training efficiency and scalability are important in choosing linear projection methods for a general face recognition problem. It has been shown that LDA/QR is much more efficient than LDA/GSVD (by theoretical and experimental analysis) [17]. Furthermore, LDA/QR is also scalable. LDA/QR is more adaptable to general face recognition problems where a straightforward solution is to choose representative instances of each unconstrained condition and combine them into a training set. Under this scenario, the efficiency can be a bottleneck of LDA/GSVD. Furthermore, if a face database is too large to let the $S_b$ fit into the memory of a computer, LDA/GSVD will fail without the scalability.

**Non-uniform behavior of linear projection methods** When integrated with multiresolution analysis, linear projection methods behave differently. This is an interesting observation in this comparative study. It is different from the results in [3] that involve the use of wavelets, but not FR/I. There are two possible factors contributing to the non-uniform behavior.

1. The use of different matrix decomposition methods in these linear projection methods. Note that both OCM and LDA/QR use QR Decomposition, whereas PCA, PCA+LDA, and LDA/GSVD use either SVD or GSVD).
2. The ways of using MRA ladder information. [3] used the low-frequency information via MRA. Here we use high-frequency information.

We tend to believe that the first factor is more likely the cause for the non-uniform behavior of linear projection methods when integrated with MRA.

It is worth mentioning here that QR Decomposition has received attention in recent face recognition under unconstrained illuminations. Both [1] and [18] use QR Decomposition to obtain othornormal bases of illumination subspaces with the motivation that it has lower time complexity than SVD. (QR Decomposition used in these works was not applied to LDA implementations.) Our work is the first one in showing that QR Decomposition can be useful not only in reducing time complexity, but also in improving recognition accuracy.

## 5 Discussions

### 5.1 Interpolation scheme

In the experiments on PIE and YaleB, we applied interpolation schemes to construct training data. By interpolation scheme, we mean that a test instance is expected to lie in the illumination subspace constructed by training instances (of some subject). There are at least two rationales for the use of interpolation scheme in linear projection methods. First, the orthogonalization of vector bases is more stable if each pair of bases vectors has large angles. Second, from the view of within-class distance, extreme training instances can achieve stronger ability to describe the structure of within-class scatters. Note that we split the data PIE/Pose27 into n pieces according to the naming locality, so reverse n-fold validation is an interpolation scheme in constructing training sets. Compared with n-fold validation, reverse n-fold validation is more challenging (and contributes better validation on a recognition technique). This is because it uses less data to train, and more data to test. Finally, the high recognition accuracy rate via reverse n-fold validation suggests that it is not necessary to use the same illumination subspaces for different subjects, which brings in flexibility of constructing training data in real-life face recognition applications.

### 5.2 How about longer wavelets?

Haar is the shortest (and simplest) wavelet. It is natural to ask the effectiveness of longer wavelets (say from the Daubechies wavelet family) in FR/I problems. We have found that the effectiveness of MRA decreases with the length of the involved Daubechies wavelets. It is a surprising result at first sight. From the view of image processing (e.g. image compression), it has been shown that longer Daubechies wavelets (say Db6) are superior to the Haar wavelet since the former has higher vanishing moments, and can more compactly represent the local information. However, the longer wavelets have a more severe shift-variance phenomenon. By shift-variant, we mean that the wavelet coefficients of a shifted function/image are different from the ones of the original function/image. It is this shift-variance property that caused the poor performance of longer wavelets.

## 6 Conclusions

In this paper, we presented a comparative study of linear projection methods in face recognition under unconstrained illuminations. It further involves the use of multiresolution analysis that surprisingly enriches the behavior of the linear projection methods. This study leads to several novel observations: i). the superiority of LDA/GSVD; ii). the tradeoff between LDA/GSVD and LDA/QR; iii). the superiority of OCM over PCA. These observations are expected to be useful when we attempt to apply the current linear projection methods to other face recognition problems.

## Acknowledgements

We would like to thank Bill Ulrich of Video/Image Modeling and Synthesis (VIMS) Lab, University of Delaware, for his helpful discussions and comments.

## References

[1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE TPAMI*, 25(2):218–233, 2003.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In *ECCV (1)*, pages 45–58, 1996.

[3] J. Chien and C. Wu. Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE TPAMI*, 24(22):1644–1649, 2002.

[4] I. Daubechies. Ten lectures on wavelets. SIAM, Philadelphia, 1992.

[5] R. Epstein, P. Hallinan, and A. Yuille. $5 \pm 2$ eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *IEEE workshop on physics-based modeling in computer vision*, pages 108–116, 1994.

[6] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE TPAMI*, 23(6):643–660, 2001.

[7] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *CVPR*, pages 995–999, 1994.

[8] K. Lee, J. Ho, and D. Kriegman. Ninepoints of light: Acquiring subspaces for face recognition under variable lighting. In *IEEE Conf. CVPR*, pages 519–526, 2001.

[9] S. Mallat. A wavelet tour of signal processing. Academic Press, 1998.

[10] H. Park, P. Howland, and M. Jeon. Cluster structure preserving dimension reduction based on the generalized singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 25(1):165–179, 2003.

[11] H. Park, M. Jeon, and J. Rosen. Lower dimensional representation of text data based on centroids and least squares. *BIT*, 43(2):1–22, 2003.

[12] R. Ramamoorthi. Analytic PCA construction for theoretical analysis of lighting variability, including attached shadows, in a single image of a convex lambertian object. *IEEE TPAMI*, 24(10):1322–1333, 2002.

[13] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. In *Proc. 4th Intl. Conf. on FG'02*.

[14] D. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE TPAMI*, 18(8):831–836, 1996.

[15] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[16] J. Ye, R. Janardan, C. park, and H. Park. An optimization criterion for generalized discriminant analysis on undersampled problems. *IEEE TPAMI, accepted*.

[17] J. Ye and Q. Li. LDA/QR: An efficient and effective dimension reduction algorithm and its theoretical foundation. *Pattern Recognition*, 37(4):851–854, 2004.

[18] L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *IEEE Conf. CVPR*, volume I, pages 19–25, 2003.

[19] W. Zhao, A. Krishnaswamy, R. Chellappa, D. Swets, and J. Weng. Discriminant analysis of principal components for face recognition. In *Wechsler, Phillips, Bruce, Fogelman-Soulie, and Huang, editors, Face Recognition: From Theory to Applications*, pages 73–85 Springer– Verlag, 1998.