Information Retrieval Using Local Linear PCA

X.Q. Li and I. King

Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
Email: {xqli, king}@cse.cuhk.edu.hk

Abstract

Efficient and accurate Information Retrieval (IR) is one of the main issues in multimedia databases. Clustering can help to generate the efficient indexing structures and provide the comparison between data types. The Most Expressive Feature (MEF) extraction can improve comparison accuracy between two data which belong to a same data type since it discards redundant features. In this paper, we introduce a Local Linear Principal Component Analysis (LLPCA) to design an optimal scheme for IR. The LLPCA realizes the clustering and local MEF extraction at the same time. Using these clusters and local MEFs, an IR scheme can be divided into two steps from coarse to fine. We apply the scheme to a trademark retrieval system to evaluate its performance based on the accuracy and efficiency measurements. The experimental results indicate this retrieval scheme is superior the other schemes using the original features or global MEFs extracted by a Global Linear PCA (GLPCA).

1 Introduction

Efficient and accurate information retrieval is an essential issue in multimedia databases. There is a well-known problem in Information Retrieval (IR) which is called "the curse of dimensionality"- more features do not necessarily imply a better retrieval accuracy rate as in pattern recognition [1, 2]. Although we often tend to use many features to describe an object perfectly, there are only a few of them can express the major aspect of the object. This leads to the selection of suitable features for accurate retrieval.

In this paper, we discuss a Local Linear Principal Component Analysis (LLPCA) to produce a set of local Most Expressive Features (MEFs) for a local distribution. The LLPCA combines the two approaches, MEF selection and clustering, together and has some good properties for efficient IR implementations.

Recent researches have shown that the retrieval approaches using only one or two "visually salient features" of a large set of features sometimes has more accurate results than using all [3]. Thus, the selection of good features is an important issue to consider. The Global Linear PCA (GLPCA) representation, also known as the Karhunen-Loève projection [5], is a wildly used technique for the feature selection or reduction. This approach projects n original features into a subspace of s features associated with the s largest eigenvalues of the sample scatter matrix, where $s < n$. Because the s produced features capture the major variations in the database, they can express the data well and can approximate the samples, where the reconstruction is very close to the original. Hence, we call the projected features as the most expressive features. Using MEFs has a number of advantages, for example, it is scale-invariant, and we can capture the essential difference of two data if they are compared in the MEF space. The neural network implementations of PCA have been introduced by Oja for subspace decomposition [6, 7].

Despite its widespread use, the GLPCA is not suitable for IR problems since there are generally many types of images in a database and different type of images perhaps need to be described with different most expressive features. For example, features for retrieving a landscape are probably useless for an animal photo. To solve the problem, we use the LLPCA [8, 11] to estimate the data types and the local MEFs of each type in our research. This LLPCA employs an extended K-means clustering algorithm to partition a data distribution into a number of non-overlapping clusters and each cluster is represented not by its central point as in clustering but by a localized linear subspace. Using LLPCA, the two important phases, the local MEF estimation and data clustering, are actually combined. Based on the LLPCA results, we divide an information retrieval approach into two steps from coarse to fine: (I) Retrieve the relevant clusters for a query according to the reconstruction distance. (II) Project the query into the MEF space of a relevant cluster. In addition, in the MEF space, retrieve the relevant data according to the Euclidean distance.

In the next section, we will describe the LLPCA in detail. Section 3 introduces an optimal IR scheme based on LLPCA. We present some experiment results to evaluate the performance of the LLPCA IR scheme by applying it to a trademark retrieval system and comparing it to the other schemes in Section 4. Last, we give a briefly discussion and conclusion in the section 5.

2 Local Linear PCA

2.1 PCA and Most Expressive Feature Extraction

In information retrieval, we can represent each data in a database using an n-dimensional feature vector $\vec{x}$, which is considered as a sample point in the n-dimensional space. The classic technique for linear dimension selection or reduction is global linear principal component analysis. In the GLPCA, one performs an orthogonal transformation to the basis of correlation eigenvectors, and projects onto the s-dimensional subspace spanned by those eigenvectors $\vec{p}_1, \vec{p}_2, \cdots, \vec{p}_s$, which correspond to the largest eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_s$, where $s < n$. The s principal components of an n-dimensional original feature vector $\vec{x}$ can be simply written as

$$\vec{y} = \mathbf{P}^T \vec{x}, \tag{1}$$

where $\mathbf{P}$ is an $n \times s$ matrix whose i-th column is the i-th principal eigenvector. The corresponding reconstruction

$$\hat{\vec{x}} = \mathbf{P}\vec{y} \tag{2}$$

will result in minimal reconstruction error in least mean square sense [4]. Hence, we consider the feature vector $\vec{y}$ consists of a set of the most expressive features of a data and the principal component representation plays the role of MEF extractor [9] [10].

2.2 Local Linear PCA

Geometrically, GLPCA models the data as a hyperplane embedded in the feature space. If the data components have non-linear dependencies, GLPCA will require a larger-dimensional representation than would be found by a non-linear technique. Kambhatla [11] proposed a locally linear approach to nonlinear dimension reduction. His algorithm first partitions the data space into disjoint regions by clustering, and then performs local PCA about each cluster. This clustering algorithm can be seen as an extended K-means algorithm in which each cluster is represented not by its central point but by a localized linear subspace. The algorithm employs the reconstruction error between a vector $\vec{x}$ and its reconstruction in the subspace as distance function called reconstruction distance instead of the Euclidean distance. Fig. 1 illustrates the difference between Euclidean distance and the reconstruction distance in a 2-dimensional space. Suppose we want to determine which of the two regions the data point $\vec{x}$ belongs. For Euclidean clustering, the distance between the point $\vec{x}$ and the two centroids $\vec{m}_1$ and $\vec{m}_2$ is compared, and the data point assigned to the cluster whose centroid is the closest; in this case is region 1. For clustering by the reconstruction distance, the distance from the point to the two 1-dimensional subspaces (corresponding to the principal subspace for the two regions) is compared, and the data point assigned to the region whose principal subspace is closest; in this case, region 2. Thus, the membership in regions defined by the reconstruction distance can be different from that defined by Euclidean distance. This is because the reconstruction distance does not count the distance along the leading eigen-directions. Neglecting the distance along the leading eigenvectors is exactly what is required, since we retain all the information in the leading directions during the PCA projection. The extended K-means algorithm is given as follows:

1. **Initialization** - Equally partition the N n-dimensional objects in database $\mathcal{X} = \{\vec{x}_k\}_{k=1}^N$ into M subsets $\mathcal{X}_1, \mathcal{X}_2, \cdots \mathcal{X}_M$. The data number in $\mathcal{X}_i$, $N_i \approx \frac{N}{M}$.

2. **Reconstruction Distance Computation** - Estimate the covariance matrices of all regions,

$$\mathbf{\Sigma}_i = \frac{1}{N_i} \sum_{\vec{x} \in \mathcal{X}_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T, \tag{3}$$

   where the $\vec{m}_i$ is the mean of i-th cluster and N_i is the data number in $\mathcal{X}_i$. Using $\mathbf{\Sigma}_i$, calculate the local eigenvectors, $\vec{v}_{i,j}, j = 1, 2, \cdots, n, \lambda_{i,1} \geq \lambda_{i,2} \cdots \geq \lambda_{i,n}$, of i-th cluster, where $\lambda_{i,j}$ is the eigenvalue corresponding to $\vec{v}_{i,j}$. Next, divide the local eigenvectors into two matrices, the principle component matrix $\mathbf{P}_i = [\vec{v}_{i,1}, \vec{v}_{i,2}, \cdots, \vec{v}_{i,s}]$ and minor component matrix $\mathbf{W}_i = [\vec{v}_{i,s+1}, \vec{v}_{i,s+2}, \cdots, \vec{v}_{i,n}]$.

   The reconstruction error J for a feature vector $\vec{x}$ in $\mathcal{X}_i$ is the reconstruction square distance between $\vec{x}$ and the local MEF subspace of $\mathcal{X}_i$,

$$J(\vec{x}, \hat{\vec{x}}_i) = (\vec{x} - \vec{m}_i)^T \mathbf{W}_i^T \mathbf{W}_i (\vec{x} - \vec{m}_i), \tag{4}$$

   where $\vec{m}_i$ is the mean of $\mathcal{X}_i$ and $\hat{\vec{x}}_i$ is the reconstruction of $\vec{x}$ calculated using Eqn.(2) and the local MEFs of i-th cluster.

3. **Repartition** - Repartition the data into M regions $\mathcal{X}_1^{new}, \mathcal{X}_2^{new}, \cdots, \mathcal{X}_M^{new}$, where

$$\mathcal{X}_i^{new} = \left\{ \vec{x} | J(\vec{x}, \hat{\vec{x}}_i) = \min_{j=1,\cdots,M} J(\vec{x}, \hat{\vec{x}}_j) \right\}. \tag{5}$$

4. **Iteration** - Iterate steps 2 and 3 until no data in each region can be assigned into the other regions.

After clustering, the LLPCA provides a set of clusters $\{\mathcal{X}_i\}_{i=1}^M$, which are represented by the means $\{\vec{m}_i\}_{i=1}^M$ and local MEF subspaces. According to the LLPCA results, we design an optimal information retrieval scheme in the next section.

3 Information Retrieval Using LLPCA

3.1 Retrieval from Coarse to Fine

Our approach retrieves the relevant data corresponding to a query $\vec{x}_q$ using the following two steps from

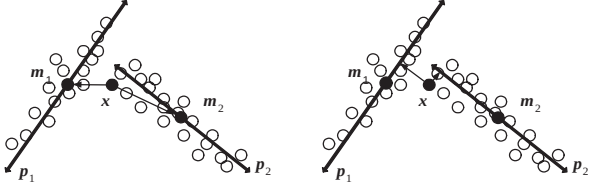


Figure 1: Assignment of the data point $\vec{x}$ to one of two regions based on (left) Euclidean distance, and (right) the reconstruction distance. The mean $\vec{m}_i$ and leading eigenvector $\vec{v}_i$ are shown for each of two regions ($i = 1, 2$).

coarse scale to fine:

1. **Cluster Retrieval** - Calculate the reconstruction distances between $\vec{x}_q$ and the MEF subspace of each cluster $\mathcal{X}_i$. The relevant clusters are ranked by ascending order according to the reconstruction distances. This step can be seen as a coarse retrieval since each retrieved cluster is a set of data.

2. **Local Data Retrieval** - Project the $\vec{x}_q$ onto the local MEF subspace of the retrieved cluster $\mathcal{X}_i$. The projection

$$\vec{y}_{i,q} = (\mathbf{P}_i)^T \vec{x}_q, \tag{6}$$

where $\mathbf{P}_i$ is an $n \times s$ matrix, which consists of s principle eigenvectors and $y_{i,q}$ is a s-dimensional vector. Compare the Euclidean square distance in the local MEF subspace between $y_{i,q}$ and each data in the retrieved cluster. The distance is

$$D(\vec{y}_{i,q}, \vec{y}_{i,j}) = (\vec{y}_{i,q} - \vec{y}_{i,j})^T (\vec{y}_{i,q} - \vec{y}_{i,j}), \tag{7}$$

where $\vec{y}_{i,j}, j = 1, 2, \cdots, N_i$ is the local MEF subspace projection of j-th data $\vec{x}_{i,j}$ in the i-th cluster. Then, the predefined number of nearest data are retrieved according to their MEF distance from the i-th cluster. This can be seen as a fine retrieval step.

3.2 Information Retrieval Scheme

Suppose we need to retrieve Z most relevant data of a query $\vec{x}_i$ in the database $\mathcal{X} = \{\vec{x}_k\}_{k=1}^N$ which has been partition into M clusters $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_M$ and the Z data perhaps are not in one cluster, it is important to design an optimal searching scheme by which only a few of clusters are visited instead of the whole database. The previous section indicates the dissimilar degree between a query $\vec{x}_q$ and the j-th data $\vec{x}_{i,j}$ in the i-th cluster depends on two distances, the reconstruction distance $J(\vec{x}_{i,q}, \hat{x}_{i,q})$ and the local MEF distance $D(\vec{y}_{i,q}, \vec{y}_{i,j})$. Using the two distances, the combined dissimilarity function of $\vec{x}_q$ and $\vec{x}_{i,j}$ can be written as,

$$E(\vec{x}_q, \vec{x}_{i,j}) = J(\vec{x}_q, \vec{\hat{x}}_{i,q}) + D(\vec{y}_{i,q}, \vec{y}_{i,j}), \tag{8}$$

where $\hat{x}_{i,q}$ is the reconstruction of $\vec{x}_q$ using the local MEFs of the i-th cluster, $\vec{y}_{i,q}$ and $\vec{y}_{i,j}$ are the projections of $\vec{x}_q$ and $\vec{x}_{i,j}$ in the local MEF subspace of the i-th cluster.

Let us consider a database which has two clusters, $\mathcal{X}_1$ and $\mathcal{X}_2$. The reconstruction distances between the query $\vec{x}_q$ and the two clusters are J_1 and J_2 and the combined dissimilarity degrees between $\vec{x}_q$ and a data $\vec{x}_{1,a}$ in $\mathcal{X}_1$ and a data $\vec{x}_{2,b}$ in $\mathcal{X}_2$ are,

$$E_1 = J(\vec{x}_q, \vec{\hat{x}}_{1,q}) + D(\vec{y}_{1,q}, \vec{y}_{1,a})$$

and

$$E_2 = J(\vec{x}_q, \vec{\hat{x}}_{2,q}) + D(\vec{y}_{2,q}, \vec{y}_{2,b})$$

From the two equations, if

$$E_1 \leq J(\vec{x}_q, \vec{\hat{x}}_{2,q}), \tag{9}$$

then $E_1 \leq E_2$. Hence, no data in cluster $\mathcal{X}_2$ need to be visited if $\mathcal{X}_1$ contains Z data and each of them satisfies Eqn. (9). This principle leads to an optimal retrieval algorithm as follows for retrieving Z data from a database, which has been partitioned into M clusters:

Step 1 Initialize the retrieved data set $\mathcal{S} = \phi$. Sort the M clusters according to J_i which is the reconstruction distance between $\vec{x}_q$ and the local MEF subspace of the i-th cluster, $\{\mathcal{X}_i\}_{i=1}^M, J_1 \leq J_2 \leq \cdots \leq J_M$. Set the first cluster to be the current cluster.

Step 2 Project $\vec{x}_q$ onto the local MEF subspace corresponding to the current cluster to calculate the combined dissimilarity between $\vec{x}_q$ and each data in the current cluster.

Step 3 A data will be put into the retrieved data set $\mathcal{S}$ if (1) $\mathcal{S} = \phi$ or (2) the dissimilarity between the data and $\vec{x}_q$ is less than the dissimilarity between a member of $\mathcal{S}$ and $\vec{x}_q$. A member will be remove from $\mathcal{S}$ if (1) the dissimilarity between it and $\vec{x}_q$ is the biggest in $\mathcal{S}$ and (2) the member number of $\mathcal{S}$ is greater than Z.

Step 4 The algorithm stops if (1) the biggest dissimilarity between a member in $\mathcal{S}$ and $\vec{x}_q$ is less than the reconstruction distance between $\vec{x}_q$ and the local MEF subspace of the next cluster, and (2) the member number is equal to Z. Otherwise, set the next cluster to current and repeat Step 2 to 4.

We evaluated the performance of the above algorithm based on precision and efficiency in the next section.

4 Experiments

4.1 Objective Criteria for Evaluation

In the experiments, we use an "artificial" trademark database to evaluate the performance of the proposed

Trademark Type	λ_1 (%)	λ_2 (%)	λ_3 (%)	Sum (%)
1	41.0294	32.1054	9.0856	82.2205
2	46.0647	24.6077	11.6204	82.2927
3	50.6555	26.4410	6.8883	83.9847
4	50.0704	28.9488	10.3310	89.3502
5	62.8324	27.6023	3.4915	93.9262
6	47.5377	27.1639	8.8805	83.5821
7	50.1863	26.9498	7.4070	84.5431
8	46.6594	22.6826	13.3207	82.6628
Global	28.7323	17.7803	9.8884	56.4010

Table 1: The percentage of the largest three eigenvalues of the total 121 eigenvalues, which are gained through PCA for each type of trademarks and all trademarks. The last column shows the sum percentages of the largest three eigenvalues.

LLCPA scheme. The "artificial" trademark database consists of 8 different types of trademark. All data in a type are generated from a "seed" trademark using different geometric distortion transformations. The distortions include Pinch, Twirl, Ripple, Horizontal Extension and Vertical Extension with 2 different degrees in each distortion as Fig. 3. Each distorted trademark then was rotated, scaled and noised to generate more data so that each type contains 39 trademarks and the whole database contains 312 data. This kind of objective evaluation criteria also has been used by Kim [3]. Based on the "artificial" trademark database , we compared the performance between the IR schemes using the original features, global MEFs and local MEFs. Here, the original features of a trademark are 121 Zernike magnitude moments [12]. To represent the distribution property of the features, we respectively do PCA for each of the 8 types of trademarks and for all trademarks in the database. Table 1 shows the percentage of the largest three eigenvalues of the total 121 eigenvalues. The last column shows the total percentages of the largest three eigenvalues and the last row shows the results of the global PCA for all data. These PCA results indicate that although the Zernike magnitude moments are wildly used features for shape description, they are highly correlated and can be reduced to a low dimensional feature space. Furthermore, we find the single type of data has greater percentages of the largest three eigenvalues than the whole database. This indicates that it is possible to use local PCA for gaining more effective feature reduction than global PCA. Next, we evaluate the retrieval schemes using the trademark database based on retrieval precision and computational efficiency.

In IR systems, precision is often defined as:

$$\text{Precision} = \frac{\text{Number of target images retrieved}}{\text{Number of images retrieved}}. \quad (10)$$

In our experiments, we simply set the goal number of trademarks retrieved equals to the number of data in each type, 39, for statistical convenience. We evaluate the computational efficiency of a retrieval method based on the two aspects: average number of visited data and average square calculation amount in each searching.

We use a set of total testing 88 queries for each experiment. This set consists of 8 sub-sets of 11 data form each data type. Each sub-set of query data includes a "seed" and 10 different distorted data. The 88 queries are shown in Fig. 2.

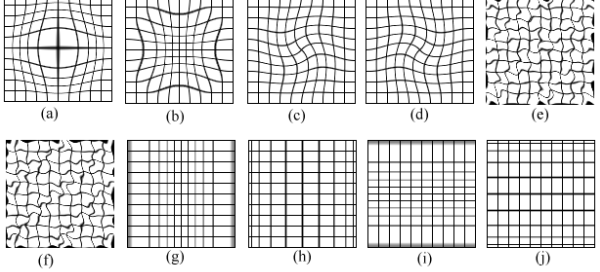


Figure 3: Distortions including Pinch: (a) and (b); Twirl: (c) and (d), Ripple: (e) and (f); Horizontal Extension: (g) and (h); and Vertical Extension: (i) and (j).

Trademark Type	Precision	NVD	SCA
1	23.31%	312	37752
2	28.21%	312	37752
3	23.08%	312	37752
4	27.97%	312	37752
5	27.04%	312	37752
6	27.27%	312	37752
7	26.81%	312	37752
8	27.74%	312	37752
Total Average	26.43%	312	37752

Table 2: The results of the IR scheme using the original 121 features and Euclidean distance.

4.2 Comparisons

We compare the results using three different retrieval schemes: (1) the LLPCA scheme using 3 local MEFs and 8 clusters, (2) the scheme using 121 original features and (3) the scheme using 3 global MEFs, where the (1), LLPCA scheme, uses the combination of reconstruction and local MEF distance, and the (2) and (3) schemes use the Euclidean distance as their similarity measures. Table 2 - Table 4 show their average retrieval precision percentage, Number of Visited Data (NVD) and Square Computation Amount (SCA) for each type of data and the total averages using the 88 testing queries.

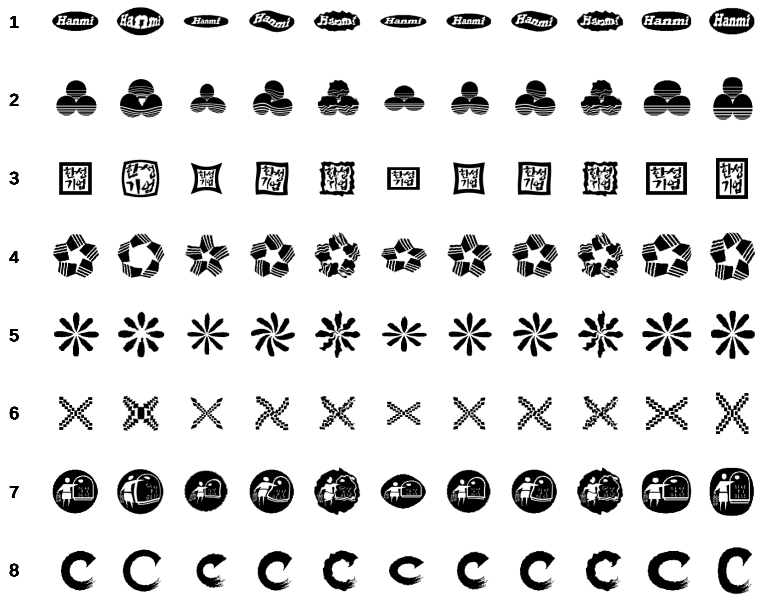


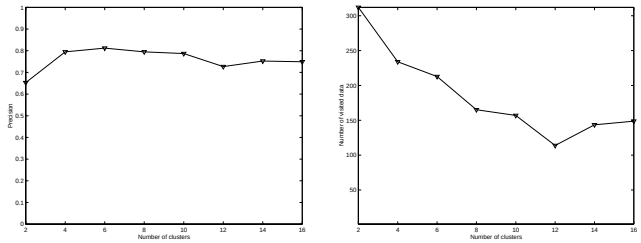
Figure 2: 88 queries from 8 types of trademarks for performance evaluation.

Trademark Type	Precision	NVD	SCA
1	22.38%	312	936
2	28.61%	312	936
3	22.14%	312	936
4	26.57%	312	936
5	26.57%	312	936
6	28.21%	312	936
7	26.11%	312	936
8	27.04%	312	936
Total Average	25.20%	312	936

Table 3: The results of the IR scheme using the 3 global MEFs and Euclidean distance.

Trademark Type	Precision	NVD	SCA
1	66.90%	146.6364	439.9091
2	96.97%	69.0909	207.2727
3	62.00%	281.5455	844.6364
4	81.59%	135.4545	406.3636
5	82.05%	252.2727	756.8182
6	89.28%	91.0000	273.0000
7	83.45%	178.7273	536.1818
8	73.66%	218.0000	654.0000
Total Average	79.49 %	171.5909	514.7727

Table 4: The results of the IR scheme based on the LLPCA. The database are partitioned into 8 clusters and each cluster use 3 local MEFs.



(a) (b)

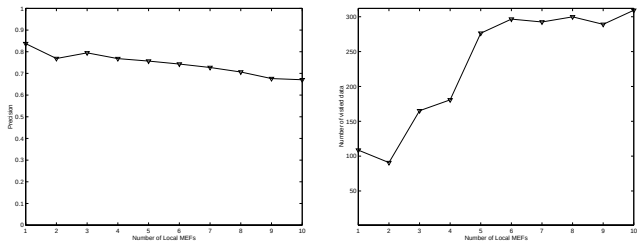
Figure 4: (a) Precision and (b) number of visited data using different number of clusters.

4.3 Number of Clusters and MEFs

Since the choice of the number of clusters and MEFs is very subjective, we evaluate the results using different number of clusters and local MEFs here. Fig. 4(a) and Fig. 4(b) show the precision percentage and number of visited data using different number of clusters. Fig. 5(a) and Fig. 5(b) show the precision and number of visited Data using different number of local MEFs.

5 Discussion and Conclusion

In this paper, we presented a new content-based retrieval approach using local MEFs extracted by the LLPCA. The approach separates an IR scheme into two sub-retrievals from coarse to fine and shows two advantages: (1) generate an optimized order for visiting the fewest number of data in each searching and (2) increase the retrieval precision through local feature selection. The experiments show our approach is superior than the others using original features or



(a) (b)

Figure 5: (a) Precision and (b) number of visited data using different number of local MEFs.

Measures	LMEFs	GMEFs	OFs
Precision	highest	lowest	middle
NVD	least	most	most
SCA	least	middle	most

Table 5: Comparison of the performance using three types of features, the Local MEFs (LMEFs), the schemes using Global MEFs (GMFEs) and Original Features (OFs) for trademark retrieval from the testing database.

global MEFs (see Table 5 for summary).

Since the features to describe an object sometimes are highly correlated especially when the number of features is large, it is important to select a set of MEFs to capture the major characteristics of an object and ignore the minor details. The MEFs usually are different to describe different data type. Hence, some local analysis approaches are necessary to estimate the most expressive features. We use LLPCA in our research. Although the LLPCA approach shows some excellent performance in our experiments, there still exist some limitations. The main problem is the parameter selection: Unsuitable choices of the number of local MEFs or clusters will decrease the performance of the LLPCA retrieval approach.

The choice of number of local MEFs mainly depends on the data distributions in feature space. We need to ensure the condition that the MEFs need occupy the very large eigenvalues by comparison to the minor features. This means the approach will not be suitable if the distribution shape of each data type is nearly hyper-spherical. Also if the features are described using different scales, we need to apply some dimension-scaling techniques before using the LLPCA approach. Hence, some supervising methods will be considered in our future research.

The choice of number of clusters is an old but still unsolved problem in practice although there have been some criteria for finding the "natural" clusters in theory. However, a type of Hierarchical Local Linear PCA algorithms [13] shows some promise to avoid the problem. The Hierarchical Local Linear PCA uses a sequential binary partition algorithm, which divides the database into 2^M subsets. A similar clustering method has been used by King [15] to generate a binary tree indexing structure for information retrieval, although [15] did not use the local MEFs as the representation of a cluster. Moreover, Lau and King indicated in [14] that the binary clustering method is fairly stable especially when the number of natural clusters in the input distribution is quite large. In our future research, we plan to use the Hierarchical Local Linear PCA to generate efficient tree structures for information retrieval.

References

[1] D. L. Swets and J. J. Weng: Using Discriminant Eigenfeatures for Image Retrieval. In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 18, No. 8, pp. 831-836, August, 1996.

[2] C.J. van Rijsbergen: Information Retrieval. Butterworths, London, U.K., 1979.

[3] Y. S. Kim and W. Y. Kim: Content-based Trademark Retrieval System using a Visually Salient Feature.

[4] I. T. Jolliffe: Principal Component Analysis. Springer-Verlag, New York, 1986.

[5] M. M. Loève: Probability Theory. Princeton, NJ: Van Nostrand, 1955.

[6] E. Oja: A Simplified Neuron Model as a Principal Component Analyzer. Journal of Mathematical Biology 15, 267-273, 1982.

[7] E. Oja: Neural Networks, Principal Components, and Subspace. International Journal of Neural Systems 1, 61-68, 1989.

[8] E. Oja: Principal Components, Minor Components, and Linear Neural Networks. Neural Networks 5, 927-936, 1992.

[9] A. Pentland, B. Moghaddam, and T. Starner: View-based and Modular Eigenspaces of Face Recognition. In Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition, pp. 84-91, June 1994. Seattle, Washington.

[10] A. Pentland, R. W. Picard, and S. Scarloff: Photobook: Tools for Content-based Manipulation of Image Databases. In SPIE Storage and Retrieval Image and Video Databases II, no. 2185, (San Jose, February 1994.

[11] N. Kambhatla and T. K. Leen: Dimension Reduction by Local PCA. Neural Computation, 9, 1493, 1997.

[12] R. R. Bailey and M. Srinath: Orthogonal Moment Features for Use with Parametric and Non-Parametric Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(4): 389-399, April 1996.

[13] L. Xu: Vector Quantization by Local and Hierarchical LMSER. Proc. of 1995 Intl Conf. on Artificial Neural Networks, Paris, France, Oct.9-13, 1995, Vol.II, 575-579.

[14] T. K. Lau and I. King: Performance analysis of clustering algorithms for information retrieval in image databases. In Proceedings to the International Joint Conference on Neural Networks (IJCNN'98), pages 932–937. IEEE Computer Press, May 4-9, 1998.

[15] I. King, L. Xu, and L.W. Chan: Using Rival Penalized Competitive Clustering for Feature Indexing in Hong Hong's Textile and Fashion Image Database. In Proceedings to the International Joint Conference on Neural Networks (IJCNN'98), pages 237–240. IEEE Computer Press, May 4-9, (1998).