Subspaces Indexing Model on Grassmann Manifold for Image Search

Xinchao Wang, Zhu Li, *Senior Member, IEEE*, and Dacheng Tao, *Member, IEEE*

***Abstract*—Conventional linear subspace learning methods like principal component analysis (PCA), linear discriminant analysis (LDA) derive subspaces from the whole data set. These approaches have limitations in the sense that they are linear while the data distribution we are trying to model is typically nonlinear. Moreover, these algorithms fail to incorporate local variations of the intrinsic sample distribution manifold. Therefore, these algorithms are ineffective when applied on large scale datasets. Kernel versions of these approaches can alleviate the problem to certain degree but face a serious computational challenge when data set is large, where the computing involves Eigen/QP problems of size N × N. When N is large, kernel versions are not computationally practical. To tackle the aforementioned problems and improve recognition/searching performance, especially on large scale image datasets, we propose a novel local subspace indexing model for image search termed Subspace Indexing Model on Grassmann Manifold (SIM-GM). SIM-GM partitions the global space into local patches with a hierarchical structure; the global model is, therefore, approximated by piece-wise linear local subspace models. By further applying the Grassmann manifold distance, SIM-GM is able to organize localized models into a hierarchy of indexed structure, and allow fast query selection of the optimal ones for classification. Our proposed SIM-GM enjoys a number of merits: 1) it is able to deal with a large number of training samples efficiently; 2) it is a query-driven approach, i.e., it is able to return an effective local space model, so the recognition performance could be significantly improved; 3) it is a common framework, which can incorporate many learning algorithms. Theoretical analysis and extensive experimental results confirm the validity of this model.**

***Index Terms*—Grassmann manifold (GM) distance, local learning model, query-driven approach, subspace selection.**

I. Introduction

SUBSPACE selection algorithms have been successfully used in many applications [22], [23], [26], [27], e.g., human face recognition [16], speech recognition and gait recognition [12]. Conventional algorithms, such as principal component analysis (PCA) [11], linear discriminant analysis (LDA) [14], [19], have proven their value in several applications. However, these algorithms are optimal under Gaussian assumption and their performance criterion is global [10]. Therefore, they fail to estimate the nonlinearity of the intrinsic data manifold, and ignore the local variation of the data [8]. Consequently, these algorithms result in unsatisfactory recognition performance in real world problems especially for large scale datasets. These global models prove often ineffective for search problems on large scale image datasets.

Recently, nonlinear algorithms have been proposed to alleviate this problem, e.g., kernel algorithms [7] and manifold learning algorithms [13], [25]. Kernel based algorithms apply a nonlinear kernel mapping on the original data while manifold learning algorithms exploit the intrinsic data distribution. With these algorithms, the recognition performance can be significantly improved. However, they face a serious computational challenge when the data set is large. This is because these algorithms involve a matrix decomposition problem of size $N \times N$, where N is the number of training samples. When N is large, these algorithms are computationally impractical and the solutions are unstable, i.e., these algorithms are inappropriate for searching problems on large scale image datasets.

To solve the aforementioned problem and improve classification performance especially on large scale datasets [4], [21], we propose an efficient localized indexing model, termed Subspace Indexing Model on Grassmann Manifold (SIM-GM). To improve the searching performance, in SIM-GM we apply Grassmann manifold measurements to manipulate the indexed subspaces derived from partitioning the global space. Moreover, we construct the model hierarchical tree so that our proposed model is able to return a customized local subspace in a query-driven manner.

To construct SIM-GM, first we partition the global sample space into local patches with a hierarchical tree structure for indexing [3], [4], [20], [24], wherein each node in the tree corresponds to a local sample space. We name the derived tree as "data partition tree", and utilize its nodes for learning and classification. The highest-level node in the tree corresponds to the global space, and the lowest-level node, i.e., leaf node, corresponds to the "smallest" local space.

In the data partition tree, leaf nodes embody the most "local" neighborhood information. However, they may fail to be the most discriminant level because of the following reasons: 1) when the number of levels in the data partition tree is large, each leaf node may contain insufficient number of samples, which yields a poor learning performance; 2) when a sample point lies on the boundaries of a node, i.e., a local space, leaf nodes may fail to offer the best discriminating power.

Manuscript received April 01, 2010; revised September 06, 2010 and January 17, 2011; accepted January 21, 2011. Date of publication February 14, 2011; date of current version August 19, 2011. This work was supported in part by a grant from Microsoft Research Asia and in part by Hong Kong RGC competitive grant 524509. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Bulent Sankur.

X. Wang is with the School of Computer and Communication Science, EPFL, Switzerland (e-mail: xinchao.wang@epfl.ch).

Z. Li is with Futurewei (Huawei) Technology, Bridgewater, NJ 08807 USA (e-mail: zhu.li@ieee.org).

D. Tao is with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia (e-mail: dacheng.tao@uts.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2114354

We solve this problem by incorporating the Grassmann manifold distances [1], [2] into our proposed model. With Grassmann distances, our model is able to manipulate the leaf nodes in the data partition tree automatically and build the most effective local space for classification. The model works in a bottom-up manner, i.e., it starts from the leaf nodes, merges the "nearest" nodes into a new one, and propagates upward. In this way, a new tree structure is created, wherein each node in the tree corresponds to a local space. We name the new tree as "model hierarchical tree". Further, we apply cross validation to empirically record the recognition performance on each node of the model hierarchical tree. Therefore, when a new query comes, our model provides a customized local space for classification, i.e., in a query-driven manner. We apply our model on a large multimedia dataset and compare its performance against other conventional algorithms. A significant improvement in recognition rate suggests the validity of our proposed model.

Our model enjoys a number of merits: 1) our localized indexing approach is efficient compared against other global algorithms, e.g., Linear Preserving Projection (LPP) [5], [15], especially when the number of training samples is large; 2) given a new query, our model is able to return a customized effective local space for classification. Therefore, the classification error rate can be significantly reduced; and 3) it is a common framework, which can incorporate many learning algorithms, i.e., we are able to introduce many conventional global subspace selection algorithms into this local model. The rest of this paper is organized as follows: in Section II, we present our proposed Subspace Indexing Model on Grassmann Manifold (SIM-GM), including the implementation of the data partition tree and the Grassmann manifold distance for manipulating nodes and the model hierarchical tree; in Section III, we present the experimental results by comparing our proposed model against other conventional models; Section IV gives the conclusion and outlines the further direction of our work.

II. Subspace Indexing Model on Grassmann Manifold (SIM-GM)

To improve the searching performance especially in large scale image datasets, in SIM-GM, we construct two tree structures: the data partition tree and the model hierarchical tree—the data partition tree to speed up the retrieval and the model hierarchical tree to record the most effective local space in a customized manner, i.e., as a query-driven approach. To construct the model hierarchical tree, we introduce Grassmann manifold distance to manipulate the local subspaces derived from partitioning the global space.

The learning procedure of our proposed model is summarized as follows: 1) apply subspace selection, i.e., Principal Component Analysis (PCA) on the global sample space; 2) apply the KD-tree based indexing on the obtained subspace, partition the global space into a number of local spaces, and derive the data partition tree; 3) apply the Grassmann manifold distance to measure the leaf nodes in the data partition tree, i.e., the distance between two nodes on the Grassmann manifold; 4) construct the model hierarchical tree based on the distance measure, use cross-validation to empirically record the recognition performance on each node of the model hierarchical tree; 5) when a new query comes, consult the model hierarchical tree, and return the most effective local space for recognition.

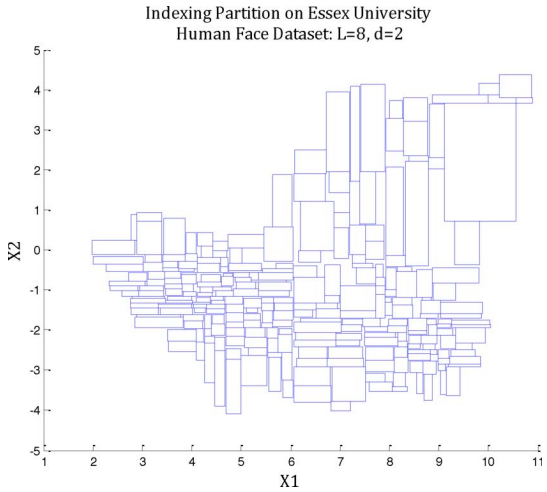


Fig. 1. Indexing space partition example on the Essex University human face dataset: $\mathrm{L} = 8, \mathrm{d} = 2$, where L is the number of levels in the data partition tree, and d is the number of dimensions preserved.

The rest of this section is organized as follows: in Section II-A we detail the implementation of the data partition tree; in Section II-B we give a review of the Grassmann manifold and incorporate it into our indexing framework; in Section II-C we present the implementation of the model hierarchical tree.

A. Data Partition Tree: A KD-Tree Based Indexing

In order to improve our indexing performance, we need samples to distribute evenly in a subspace with a low dimension, so that we could apply a KD-tree based partition scheme to divide the whole space into nonoverlapping subspaces more efficiently. Conventional subspace selection algorithms could be applied on the whole sample space before the whole space is partitioned and indexed.

PCA can be an effective approach for the indexing. It selects the first d bases with largest variance, here we denote the first d bases as $A_{\mathrm{Index}} = [a_1, a_2, \ldots a_d]$. The covariance information obtained from global PCA is utilized in the indexing. The indexing process is described as follows: 1) project all sample points on the maximum variance basis a_1, find the median value of the projected samples m_1, and split the whole collection of data along a_1 at m_1, i.e., split the current node into left and right children; 2) starting from $i = 2$, for each left and right child, project the whole collection of data along the i-th maximum variance basis a_i, find the median value m_i, and split all the children at m_i 3) increment i, repeat 2) until some predefined criteria for number of levels, or the number of samples in the leaf node is satisfied. At each node, a minimum bounding box (MBB), i.e., $V_{\min}, V_{\max} \in R^d$, is computed and recorded. The split dimension and medium values are also recorded.

We apply our indexing scheme on Essex University dataset [9] and plot Fig. 1. For the number of levels in the data partition tree, we assign $L = 8$; for the dimension of data, in order to get a better visualization, we assign $d = 2$, i.e., we apply PCA and reduce the dimension of original data to 2.

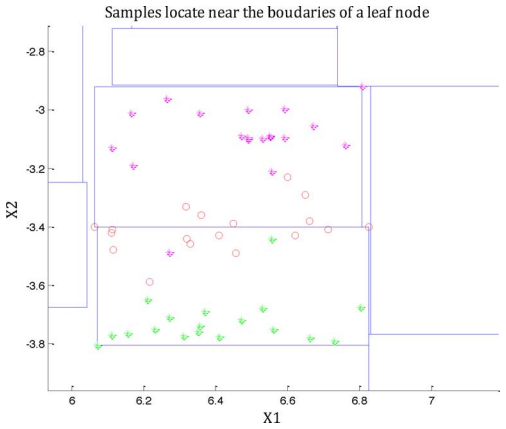


Fig. 2. Example when samples distribute near the boundaries of a leaf node.

In Fig. 1, each box corresponds to a leaf node in the data partition tree, i.e., the smallest local sample space. Sample points in the same leaf node are considered to be close in the Euclidean space, and thus, the leaf nodes could be utilized for localized discrimination.

The partition on the global space serves as the first stage of the proposed work. Compared with other local partition methods e.g., Quad-tree partition, kd-tree partition is superior because this approach is distribution-based partition and minimizes the quantization error. However, when the number of local patches is sufficient enough, the constructed local models will be the same for any given query point and locality.

However, leaf nodes may not be the most effective local space for classification, because of several reasons. First, as the number of levels in the data partition tree increases, the number of samples in each leaf node decreases. Therefore, each leaf node may contain insufficient number of samples to offer a strong discriminating power. We may consider an extreme case, if the number of levels in the data partition tree $L = \log_2 N$, where N is the total number of samples, each leaf node contains only one sample. Under such circumstance, when a new query point comes, to assign a leaf node (i.e., local space) to it is equivalent to find its nearest neighbor in the Euclidean space. This approach is obviously not what we want.

Second, the leaf nodes are ineffective for classification when the training samples lie near the node boundaries. This case indicates similarities between the two leaf nodes. This example can be visualized in Fig. 2.

Fig. 2 shows the case when samples points in the training set distribute near the boundaries of leaf nodes. Under such circumstance, there are samples from three classes. The class in the middle, i.e., the class with the symbol 'o', distribute along the boundary. When a new query of class 'o' comes, it will be assigned to either one of the two leaf nodes. However, as one may observe, the most promising local space is the merged space of the two leaf nodes.

B. Grassmann Manifold Distance for Manipulating Nodes in the Data Partition Tree

To alleviate the aforementioned problems of leaf nodes, we introduce Grassmann manifold into our indexing model for manipulating the leaf nodes derived from the data partition tree. We utilize the Grassmannian metric and build the most effective local space. Therefore, when a new query comes, our model returns the most effective local space for classification. In this section, we give a brief review on the concept of Grassmann manifold [1], [2] and the Grassmann manifold distance, i.e., principal angles and geodesic distance.

First, we briefly review the definition of Grassmann manifold. Grassmann manifold $G(d, D)$ can be defined as the set of d-dimensional linear subspace in R^D [2]. We consider the space $\mathrm{R}^{(0)}_{\mathrm{D,d}}$ of all $D \times d$ matrices, i.e., $Y \in R^{D\times d}$. The group of transformation $Y = YL$, where L is a full rank $d \times d$ square matrix, defines a equivalence relation in $\mathrm{R}^{(0)}_{\mathrm{D,d}}$

$$Y_1 = Y_2 \text{ if } \mathrm{span}(Y_1) = \mathrm{span}(Y_2) \quad \text{where } Y_1, Y_2 \in R^{(0)}_{D,d}. \tag{1}$$

Therefore, the equivalence classes of $\mathrm{R}^{(0)}_{\mathrm{D,d}}$ are in one-to-one correspondence with the points on the Grassmann manifold $G(d, D)$, i.e., each point on the manifold is a subspace.

Grassmann manifold $G(d, D)$ can be seen as a quotient space

$$G(d, D) = R^{(0)}_{D,d}/R^{(0)}_{d,d} \tag{2}$$

where the dimension of the analytical manifold $G(d, D)$ is $Dd-d^2$. When we consider Y as point in $R^{D\times d}$, the set of all elements YL in the equivalence class forms a surface of dimension d^2 in $R^{D\times d}$.

Second, we introduce principal angles and principal vectors. Each point on the Grassmann manifold is a subspace. Therefore, to measure the distance between two points on the Grassmann manifold is equivalent to measure the similarities between two subspaces. Principal angle [1], [2], [17], [18] is a geometrical measure between two subspaces, i.e., a measure of distance on the Grassmann manifold. We consider two orthonomal matrices $Y_1, Y_2 \in R^{D\times d}$ on the Grassmann manifold, the principal angles $0 \le \theta_1 \le \cdots \le \theta_d \le \pi/2$ between two subspaces $\mathrm{span}(Y_1)$ and $\mathrm{span}(Y_2)$ are recursively defined by

$$\begin{aligned} \cos\theta_k &= \max_{u_k \in \mathrm{span}(Y_1)} \max_{v_k \in \mathrm{span}(Y_2)} u_k' v_k \\ \text{s.t.} \quad & u_k' u_k = 1,\ v_k' v_k = 1 \\ & u_k' u_i = 0,\ v_k' v_i = 0 \\ & \text{for } i = 1, \ldots, k-1. \end{aligned} \tag{3}$$

The vectors $(u_1, u_2, \ldots, u_d)$ and $(v_1, v_2, \ldots, v_d)$ are called principal vectors of the two subspaces. θ_k is the kth smallest angle between two principal vectors u_k and v_k, e.g., θ_1 is the smallest principal angle and $\cos\theta_1 = u_1' v_1$.

Third, we focus on the computation of principal angles and principal vectors. Several ways exist to compute the principal angles between two subspaces. One numerically stable way is to apply Singular Value Decomposition (SVD) on the product of the two matrices $Y_1'Y_2$, i.e.,

$$Y_1'Y_2 = USV' \tag{4}$$

where $U = [u_1, u_2, \ldots, u_d], V = [v_1, v_2, \ldots, v_d]$ and $S = \mathrm{diag}(\cos\theta_1, \ldots, \cos\theta_d)$. The cosine of principal angles θ, i.e., $\cos\theta_1, \ldots, \cos\theta_d$ are known as canonical correlations [17].

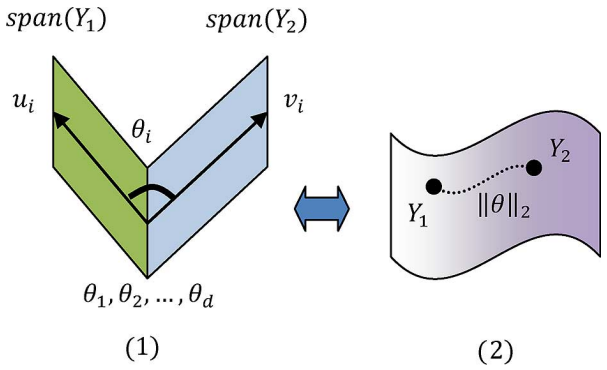


Fig. 3. Principal angles in R^D and Grassman Distance in $G(d, D)$. $\mathrm{span}(Y_1)$ and $\mathrm{span}(Y_2)$ are two subspaces of dimension d in R^D; the distance between these two subspaces can be measured by principal angles $\theta = [\theta_1, \theta_2, \ldots, \theta_d]$. In the Grassmann manifold point of view, two subspaces $\mathrm{span}(Y_1)$ and $\mathrm{span}(Y_2)$ are two points on the manifold $G(d, D)$, the geodesic distance between these two points on the manifold is $d(Y_1, Y_2) = \|\theta\|_2$. (1) principal angle in R^D, (2) Grassmann distance in $G(d, D)$.

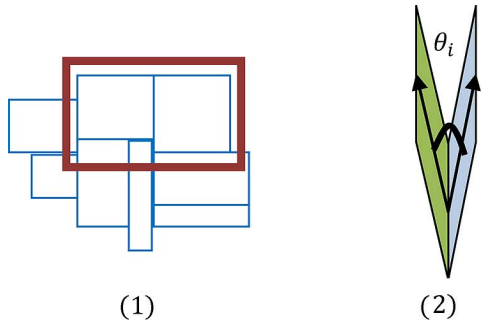


Fig. 4. Criterion for defining two "similar" nodes in the node set: First, the two nodes should be adjacent in Euclidean space; second, among all the adjacent node pairs, the two nodes (i.e., subspaces) should be of shortest distance on the Grassmann manifold, i.e., the smallest principal angles. If the two nodes satisfy the above criterion, we merge these nodes into a new one and replace the original two nodes with this new node in the node set. The construction process is repeated until there is only one node in the node set. (1) adjcent in Euclidean space, (2) close on Grassmann manifold.

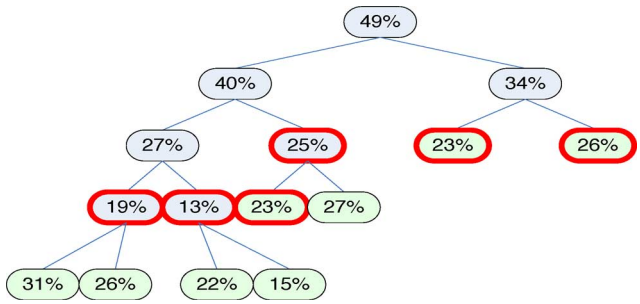


Fig. 5. Example of model hierarchical tree. The number inside each node indicates an empirical classification error rate. A bold node indicates the most effective level for classification. When a new query comes, our model will assign it to the corresponding the most effective level for recognition.

Finally, we define the distance on the Grassmann manifold. A distance is referred to as Grassmann distance if it is invariant under different basis representations. Grassmannian distances between two linear subspaces $\mathrm{span}(Y_1)$ and $\mathrm{span}(Y_2)$ can be described by principal angles. The smaller principal angles are, the more similar two subspaces are, i.e., the closer they are on the Grassmann manifold. In literature, a number of subspace distances are proposed, e.g., projection distance, Binet-Cauchy distance, and Max/Min Correlation. In this paper, we adopt the geodesic distance (or arc-length) [1], which is defined as follows:

$$d^2_{\mathrm{Arc}}(Y_1, Y_2) = \sum_i \theta_i^2. \tag{5}$$

As we may observe in Fig. 3, the geodesic distance is derived from the geometry of Grassmann manifold. It is the length of geodesic curve connecting two subspaces along the Grassmann surface. The geodesic distance decreases as the principal angles decrease; when $\theta_1 = \theta_2 = \cdots = \theta_d$, the distance between two subspaces is zero and two subspaces collapse into one.

The reasons we choose geodesic distance as the Grassmann manifold distance lie on the following aspects: 1) geodesic distance is a metric which satisfies a number of properties, e.g., symmetric property and triangular property; some other distance measures (e.g., max correlation) are not metrics so they do not have such properties [1]; 2) the max correlation is a robust distance when the subspaces are highly noisy, while the min correlation is more discriminant when data are concentrated and have nonzero intersections. Geodesic distance have intermediate characteristic so it can be applied for a wider range of data distributions.

C. Model Hierarchical Tree

As we discussed in Section II-A, the leaf nodes of the data partition tree may not be the most effective local spaces for classification. To obtain the most effective local space and improve the classification performance, we apply Grassmannian distance on the leaf nodes and measure their similarity. Then we construct a model hierarchical tree and record the recognition rate on each level. When a new query comes, we assign it to the most promising level in the model hierarchical tree for classification.

To construct the model hierarchical tree, we initialize the "node set" with leaf nodes of the data partition tree, and propagate the tree in a bottom-up manner by merging the "similar" nodes in the node set. The criterion for two nodes Y_i, Y_j to be "similar" is defined as follows: 1) in the Euclidean space, the two nodes should be adjacent, i.e., $Y_j \in \mathrm{adj}(Y_i)$; 2) on the Grassmann manifold, the two nodes should be close in geodesic distance, i.e.,

$$\begin{aligned} j &= \arg\min_k d^2_{\mathrm{Arc}}(Y_i, Y_k) \\ \text{s.t.} \quad & Y_k \in \mathrm{adj}(Y_i). \end{aligned} \tag{6}$$

We may geometrically understand the criterion as follows: the two local subspaces should be not only near to each other in space, but also similar in "shape". One example can be visualized in Fig. 4.

To merge two leaf nodes, i.e., two local subspaces, we apply the Grassmann manifold distance again. Let us consider two subspaces $\mathrm{span}(Y_1)$ and $\mathrm{span}(Y_2)$, with corresponding principal vectors $U = [u_1, u_2, \ldots, u_d], V = [v_1, v_2, \ldots, v_d]$. Because U and V are orthogonal, they can be considered as bases for the two subspaces $\mathrm{span}(Y_1)$ and $\mathrm{span}(Y_2)$ correspondingly,

i.e., $\mathrm{span}(U) = \mathrm{span}(Y_1)$ and $\mathrm{span}(V) = \mathrm{span}(Y_2)$. We manipulate the bases of the two original subspaces and create the basis for the merged subspace, i.e., we utilize U and V and create the basis T for the new subspace. The basis T can be derived as follows:

$$t_k = \frac{n_1}{n_1+n_2}u_k + \frac{n_2}{n_1+n_2}v_k \quad \text{for } k = 1, \ldots d \quad (7)$$

where n_1 is the number of samples in node 1, n_2 is the number of samples in node 2, and $T = [t_1, t_2, \ldots, t_d]$. As we may observe, the derived basis t_k lies on the hyper-plane spanned by u_k and v_k.

To prove the derived T is a set of basis for the new subspace, we have the following theorem:

Theorem 1: The derived $T \in R^{D\times d}$ from (7) is a set of basis for a subspace of dimension d in R^D.

Proof: T is the basis set for a subspace of dimension d indicates that the columns in T are linearly independent. Further, columns in T are linearly independent if and only if $T'T$ is positive definite.

For any $X \in R^{d\times 1}$

$$X'(T'T)X = (TX)'(TX) = \|TX\|^2 = \left\| \left(\frac{n_1}{n_1+n_2}U + \frac{n_2}{n_1+n_2}V \right) X \right\|^2 \geq 0. \quad (8)$$

For any $k = 0, 1, \ldots, d$, we have $\|u_k\| = 1$ and $\|v_k\| = 1$. Therefore, $X'(T'T)X = 0$ if and only if $n_1 = n_2$ and $U = -V$, i.e., $u_k = -v_k$ for all k. However, when $U = -V$, we have $\mathrm{span}(U) = \mathrm{span}(-V)$, i.e., the two subspaces collapse to one. This case will never happen because U and V are derived from $\cos\theta_k = \max\max u_k' v_k$. If two subspaces collapse into one, we would derive $U = V$ for the sake of maximizing $u_k' v_k$. Therefore, $X'(T'T)X \neq 0$, i.e., $X'(T'T)X > 0$. Further, we know columns in T are linearly independent, so T could be served as a set of basis for a subspace of dimension d. ■

However, the derived T is not orthogonal. In order to obtain the orthogonal basis for the new subspace of dimension d, we can apply Gram-Schmidt [18] process to obtain the orthogonal basis $T' = [t_1', t_d', \ldots, t_d']$.

The derived subspace can be intuitively understood as a linear "combination" of original subspaces, the linear coefficient is proportional to the number of samples in each subspace (i.e., the number of nodes in the node set).

We then replaced the two original nodes with the merged node in the node set. Our merging algorithm repeats until there is only one node in the node set, and the very last node can be considered as the global space. In this way, the model hierarchical tree is built in a bottom-up manner. The leaf nodes are those nodes obtained from the data partition tree, the inner nodes are merged nodes, and the root node corresponds to the global space.

After we set up the model hierarchical tree, we apply cross validation on our training sets and empirically derive the classification performance on each level, i.e., on each subspace. For any leaf node, there is one unique path from the root node to this leaf node; we mark the node (i.e., level in the model hierarchical tree) with the lowest classification error rate along each path. When a new query comes, we first check which leaf node it belongs to, and suggest the most promising level (i.e., the most promising local space) for classification. After we assign the incoming query to its most promising local space, we project the query and all the samples in this local space onto the subspace learnt by conventional algorithms, e.g., LDA. Then we apply KNN to classify the query. Therefore, the proposed SIM-GM is able to incorporate many learning algorithms.

The splitting operation on each local subspace is possible but not encouraged, the reasons are illustrated as follows: 1) the splitting operation forces samples from each class to be restricted within one local space (i.e., one node), which leads to problems like over-fitting; 2) it forces the using of linear classifiers, i.e., we are forced to apply linear classifier to split the original space into two, so we are not able to utilize other effective classifiers like K-NN; 3) it is more time/effort consuming.

III. Experiments

In this section, we conduct experiments on two different applications, i.e., multimedia image classification and human face recognition. To confirm the validity of our proposed model, we compare the performance of our proposed SIM-GM against those of other models. We take both computation time and classification error rate into comparison.

A. Dataset

We adopt the Microsoft Research Asia Multimedia (MSRA-MM) image dataset [6] and Essex University human face dataset [9], and compare the performance of SIM-GM against those of other models.

1) Microsoft Research Asia Multimedia (MSRA-MM) Image Dataset: The MSRA-MM consists of two sub-datasets, i.e., an image dataset and a video dataset. In this paper, we utilize the image dataset for training and testing. The image dataset contains 68 classes, each of which consists of around 1 000 images, and in total 65 443 images. All the images are collected from the query log of Microsoft Live Search. Based on the relevance, each image is assigned a relevance level: very relevant, relevant and irrelevant. These three levels are indicated by scores 2, 1, and 0, respectively.

For this image dataset, a set of features are available including: 1) 225D block-wise color moment; 2) 64D HSV color histogram; 3) 256D RGB color histogram; 4) 144D color correlogram; 5) 75D edge distribution histogram; 6) 128D wavelet texture; 7) 7D face features. Therefore, for each sample, we have 899D data in total.

As we may observe from Fig. 6, the content of the "very relevant" samples match well with their labels. The "relevant" samples are somehow redundant in content, while the "irrelevant" samples do not match their labels. To assess our proposed model, we leave out the "Irrelevant" samples in the dataset, i.e., the samples whose Relevance Indicator is 0. Therefore, we derive a dataset consisting of 52 336 samples from 68 classes. From this derived dataset, we further select images from 12 classes (i.e., in total 11555 images) and conduct classification on these images. The 12 classes include: *tree, bird, email,*

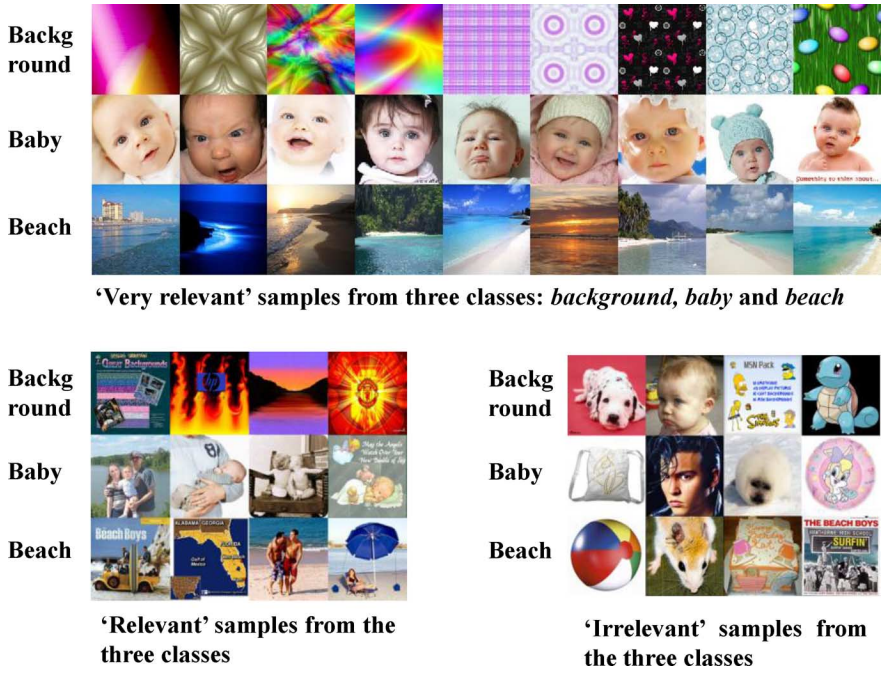


Fig. 6. Sample images from the MSRA-MM dataset.

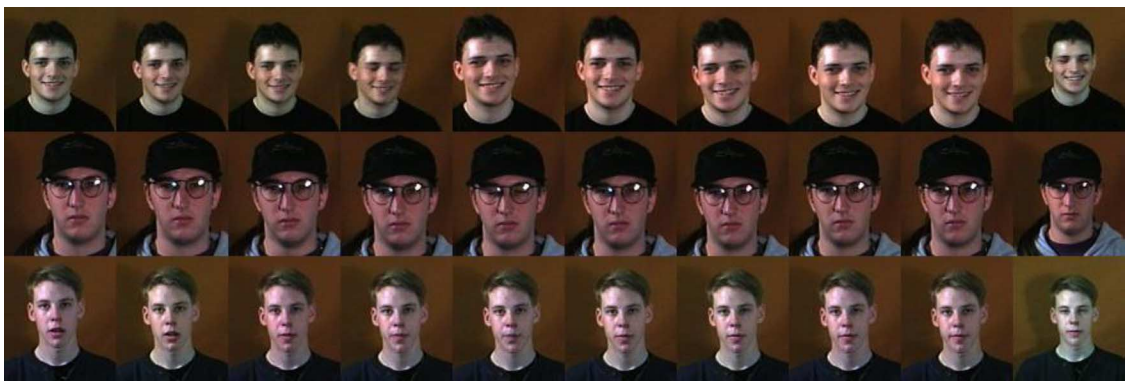


Fig. 7. Sample images from the Essex University human face dataset. Images are from three individuals.

beach, panda, youtube, military, fruit, background, dragon, people, hairstyle. For each sample, we integrate all features (i.e., 899D) for classification.

2) Essex University Human Face Dataset: The Essex University human face dataset consists of four subsets, i.e., faces94, faces95, faces96 and grimace. Images from faces96 are of size 196×196 pixels while images from the other three are of size 180×200 pixels. We assemble faces 94, faces95 and grimace subsets, and obtain a large dataset with 4 840 faces of 242 individuals, i.e., 20 faces for each individual. We then randomly select training samples and testing samples from the obtained dataset and conduct experiments.

B. General Experiments and Results

We conduct experiments on the two aforementioned datasets, and compare performance of SIM-GM against other conventional algorithms. To apply SIM-GM, we first conduct PCA on the whole dataset, and derive a subspace; then we build the data partition tree for our training data; after that we construct the model hierarchical tree based on leaf nodes derived from the data partition tree; last we apply cross validation to empirically record the most effective level for classification in the model hierarchical tree. When a new query comes, SIM-GM outputs the optimal localized subspace for classification.

We design our experiments into the following parts: 1) compare the recognition rate of SIM-GM against those of global models, e.g., global PCA; 2) compare the recognition rate of SIM-GM against those of local models; 3) compare the execution time of SIM-GM against those of global models.

We compare three different linear subspace selection approaches, i.e., PCA, LDA, and LPP. LPP [5] is a linear projection which preserves neighborhood structure of the data. It linearly approximates to the eigenfunctions of the Laplace Beltrami operator on the manifold.

1) Microsoft Research Asia Multimedia (MSRA-MM) Image Dataset: Fig. 8 shows the recognition performances of the proposed SIM-GM against those of global models. As we may observe, our SIM-GM outperforms the traditional models significantly. For the MSRA dataset, PCA preserves a better recogni-

Fig. 8. Recognition performance of SIM-GM against those of global models, i.e., PCA, LDA, and LPP.

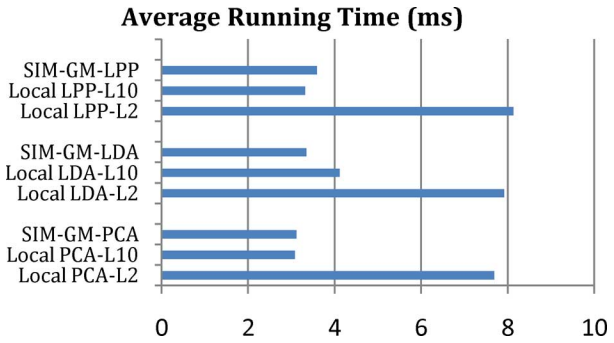


Fig. 9. Average running time for classifying one query, running on Intel Centrino VPro 2, 2 G RAM. (avg. of 10 trials, 90% Train, No. of Dimension: 30).

tion performance than LDA, while LPP outperforms PCA. This is because LDA assumes samples from each class distribute in a Gaussian distribution, which may not true for the real data; moreover, LDA can obtain a subspace of dimensionality up to C-1, where C is the number of classes; however, LPP embeds a graph distribution into the training process, so the recognition performance can be enhanced. Our SIM-GM partitions the whole sample spaces into a number of local spaces, and approximates the intrinsic data distribution by piece-wise linear local subspaces. When a new query comes, SIM-GM provides a customized effective subspace accordingly. Therefore, the recognition performance can be significantly improved. Please note that the subspaces do not necessarily have the same dimensionality in each local space. However, in order to have a fair comparison against global models, we force the dimensionality on each level to be equal.

In Table II, we compare the recognition performance of SIM-GM against those of local models. We derive the local

TABLE I
RELEVANCE LEVEL AND CORRESPONDING PERCENTAGES OF THE MSRA-MM IMAGE DATASET

Relevance Level	Relevance Indicator	Number of Images	Percentage
Very Relevant	2	19517	29.82%
Relevant	1	32819	50.15%
Irrelevant	0	13107	20.03%

TABLE II
RECOGNITION PERFORMANCE OF SIM-GM AGAINST THOSE OF LOCAL MODELS ON DIFFERENT DIMENSIONS (I.E., 5, 10, 30,AND 50) AND ON DIFFERENT LEVELS (I.E., 2, 5, 8, AND 10) IN THE DATA PARTITION TREE. (AVG. OF 10 TRAILS, 90% TRAIN). "SIM-GM-PCA/LDA/LPP" MEANS WE ASSIGN THE QUERY TO THE RECORDED MOST EFFECTIVE LOCAL SPACE IN THE MODEL HIERARCHICAL TREE, AND APPLY PCA/LDA/LPP ON THIS LOCAL SPACE FOR CLASSIFICATION. LOCAL PCA-LX MEANS WE ASSIGN THE QUERY TO A PREDEFINE LEVEL L IN THE MODEL HIERARCHICAL TREE, AND APPLY PCA ON THIS LOCAL SPACE, I.E., LEVEL L

	Recognition Performance	Running Time (ms)
PCA	0.2869	13.54
SIM-GM-PCA	0.5469	**2.63**
SIM-ES-PCA	0.3841	2.93
LDA	0.2336	14.02
SIM-GM-LDA	0.4794	2.89
SIM-ES-LDA	0.3598	3.32
LPP	0.3552	14.79
SIM-GM-LPP	**0.5786**	2.91
SIM-ES-LPP	0.3918	3.06

TABLE III
AVERAGE RECOGNITION PERFORMANCE AND RUNNING TIME FOR ONE QUERY ON AN INTEL CENTRINO VPRO 2, 2 G RAM. (AVG. OF 10 TRAILS, 60% TRAIN, NO. OF DIMENSION: 30). SIM-ES-XXX REFERS TO THE EUCLIDEAN BASED MERGING MODEL, I.E., SUBSPCE INDEXING MODEL ON EUCLIDEAN SPACE

	Recognition Performance	Running Time (ms)
PCA	0.2869	13.54
SIM-GM-PCA	0.5469	**2.63**
SIM-ES-PCA	0.3841	2.93
LDA	0.2336	14.02
SIM-GM-LDA	0.4794	2.89
SIM-ES-LDA	0.3598	3.32
LPP	0.3552	14.79
SIM-GM-LPP	**0.5786**	2.91
SIM-ES-LPP	0.3918	3.06

models from the data partition tree, and select different-level local spaces (i.e., nodes on different levels in the data partition tree) for classification. We may observe the SIM-GM is the optimal among all the local models. We may also observe the SIM-GM-LPP presents satisfactory performance. That is because LPP encodes the local geometry of training samples and finds the optimal linear embedding transformation.

Time in Table III refers to the time slot between a query arrives and an output is returned, i.e., the running time.

TABLE IV
BEST RECOGNITION RATE OF FIVE MODELS, I.E., PCA, LDA, LPP, SIM-GM-PCA, AND SIM-GM-LDA

% of Train	PCA	LDA	LPP	SIM-GM-PCA	SIM-GM-LDA
40	0.8673	0.8923	0.9087	0.9058	**0.9303**
70	0.9006	0.9484	0.9594	0.9784	**0.9987**

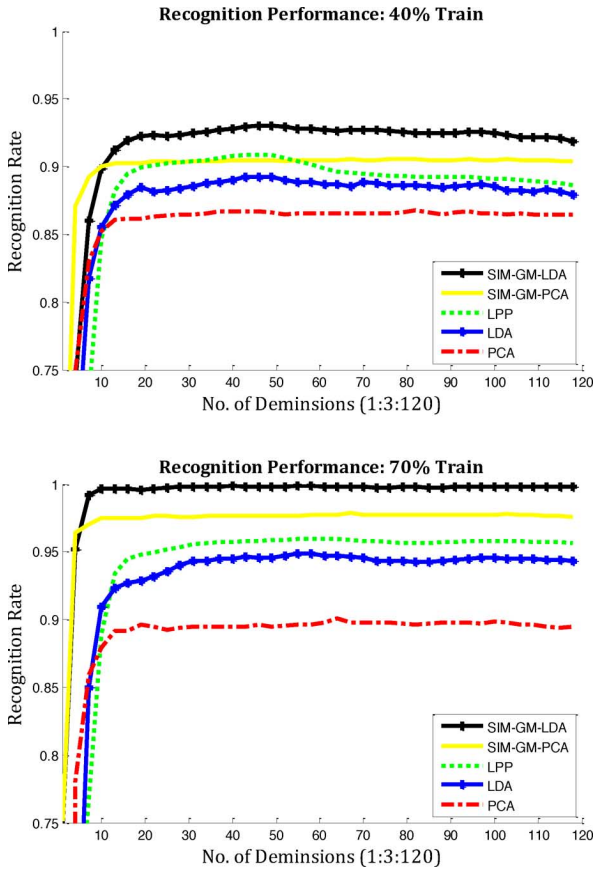


Fig. 10. Recognition rate versus number of dimensions (1:3:120). Five models are compared, i.e., PCA, LDA, LPP, SIM-GM-PCA, and SIM-GM-LDA. Every trial is repeated for ten times, and the average is recorded.

Our SIM-GM builds the model hierarchical tree and utilizes cross-validation to record the most effective level for classification in an off-line manner. SIM-ES follows similar construction process but applies different criterion to merge nodes (as compared against SIM-GM's criterion in Fig. 4): instead of using two criterions i.e., adjacent in Euclidean space and close on Grassmann manifold, SIM-ES adopts only one criterion to merge similar nodes–closest on the Euclidean space. SIM-ES computes the centroid of each node and utilizes the distance between centroids as the distance between nodes; for any node, SIM-ES finds the closest one in Euclidean space and merge them. From Table III we may observe, SIM-GM is very efficient in classification and is able to maintain a satisfactory recognition performance.

2) Essex University Human Face Dataset: In Fig. 10, we show the recognition rate versus the dimensionality. We initial the dimensionality to be 1 and increment the dimensionality by 3 each time until it reaches 120. We may observe the recognition rate trend given the change of dimensionality: the rate significantly goes up when the dimensionality is between 1 and 15; it peaks when the dimensionality is between 40 and 50; it slowly decreases/remains the same when the dimensionality is larger than 50.

TABLE V
AVERAGE RECOGNITION RATE AND RUNNING TIME FOR ONE QUERY (70% TRAIN, NO. OF DIMENSION: 40) ON AN INTEL CENTRINO VPRO 2, 2 G RAM

	Recognition Rate	Time (ms)
PCA	0.8945	14.62
SIM-GM-PCA	0.9764	**3.87**
LDA	0.9448	14.82
SIM-GM-LDA	**0.9987**	4.65

As we may observe, the recognition rates on this Essex University human face dataset are high, even for global algorithms like PCA. This is because for this dataset, the intravariance is small, as we may notice from Fig. 7. However, our proposed SIM-GM still enjoys a better performance in terms of classification rate.

Similar to MSRA-MM dataset, in Tables IV and V we show recognition rate and classification time. We may observe that our proposed SIM-GM enjoys a higher recognition rate with a shorter classification time, compared with conventional global models.

IV. CONCLUSION AND FUTURE WORK

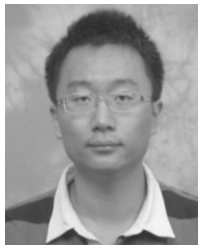
In this paper, we propose the Subspace Indexing Model on Grassmann Manifold (SIM-GM) for large subject set pattern recognition. SIM-GM partitions the manifold space into local patches with a hierarchical structure, and train local subspace models for classification. By further introducing the Grassmann manifold distance into this framework, local models are organized into model hierarchy with a second tree structure for classification at query time. The model enjoys a number of advantages including: 1) its classification efficiency on large scale image datasets; 2) its query-driven nature; 3) its ability to incorporate conventional algorithms as a framework.

In the future, we plan to introduce more learning algorithms into SIM-GM and apply this framework on other applications, e.g., document categorization.

REFERENCES

[1] J. Hamm, "Subspace-Based learning with grassmann kernels," Ph.D. dissertation, 2008.
[2] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: A unifying view on subspace-based learning," in *Proc. Int. Conf. Mach. Learning*, 2008, pp. 376–383.
[3] Z. Li, L. Gao, and A. K. Katsaggelos, "Locally embedded linear subspaces for efficient video indexing and retrieval," in *Proc. Int. Conf. Multimedia & Expo*, 2006, pp. 1765–1768.

[4] Z. Li, Y. Fu, J. Yuan, T. S. Huang, and Y. Wu, "Query driven local linear discriminant models for head pose estimation," in *Proc. Int. Conf. Multimedia & Expo*, 2007, pp. 1810–1813.
[5] X. He and P. Niyogi, "Locality preserving projections," *Neural Inf. Process. Syst.*, pp. 153–160, 2003.
[6] H. Li, M. Wang, and X. Hua, "MSRA-MM 2.0: A large-scale web multimedia dataset," in *Proc. Workshop Int. Conf. Data Min.*, 2009, pp. 164–169.
[7] J. Ham, D. D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," presented at the Int. Conf. Mach. Learning, 2004, DOI:10.1145/1015330.1015417.
[8] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.*, vol. 4, pp. 119–155, 2003.
[9] D. Hond and L. Spacek, "Distinctive descriptions for face processing," in *Proc. Brit. Mach. Vision Conf.*, 1997, pp. 320–329.
[10] V. Strassen, "Gaussian elimination is not optimal," *Numer Math.*, vol. 13, pp. 54–356, 1969.
[11] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.
[12] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[13] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 1, pp. 1–48, 2006.
[14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[15] D. Cai, X. He, J. Han, and H. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3608–3614, Nov. 2006.
[16] Y. Fu and T.-S. Huang, "Image classification using correlation tensor analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 226–234, Feb. 2008.
[17] T. Wang and P. Shi, "Kernel grassmannian distance and discriminant analysis for face recognition from image sets," *Pattern Recognit. Lett.*, vol. 30, no. 13, pp. 1161–1165, Oct. 2009.
[18] G. Strang, *Computational Science and Engineering*. Wellesley, MA: Wellesley Cambridge Press, 2007.
[19] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
[20] H. Dutağacı, B. Sankur, and Y. Yemez, "Subspace building for retrieval of general 3D models," *Comput. Vis. Image Understand.*, vol. 114, pp. 865–886, 2010.
[21] Z. Li, A. K. Katsaggelos, and B. Bandhi, "Fast video shot segmentation and retrieval based on trace geometry in principal component space," in *Proc. IEE Vision, Image Signal Process.*, May 2005, vol. 152, no. 3, pp. 367–373.
[22] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Min. Knowl. Disc.*, vol. 22, no. 3, pp. 340–371, May 2010.
[23] W. Bian and D. Tao, "Max-min distance analysis by using sequential SDP relaxation for dimension reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 1037–1050, May 2011.
[24] C. B. Akgül, B. Sankur, F. Schmitt, and Y. Yemez, "Similarity learning for 3D object retrieval using relevance feedback and risk minimization," *Int. J. Comput. Vis.*, vol. 89, pp. 392–407, 2010.
[25] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative non-negative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, 2011, DOI: 10.1109/TIP.2011.2105496.
[26] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularization for transfer subspace learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 7, pp. 929–942, Jul. 2010.
[27] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.



Xinchao Wang received the first honorable degree from the Hong Kong Polytechnic University (HK-PolyU) in 2010. He is currently pursuing the Ph.D. degree at the School of Computer and Communication Science, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

From May to September 2009, he was a Research Assistant at Nanyang Technological University, Singapore. His current research interests include applied machine learning, image indexing/retrieval, and computer vision.

Mr. Wang holds the GPA record in the Department of Computing, HK-PolyU. He is a recipient of Hong Kong Government Scholarship, i.e., the highest level scholarship in Hong Kong.

Zhu Li (SM'07) received his Ph.D. degree in electrical and computer engineering from Northwestern University, Evanston, IL, in 2004.

He was an Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University, from 2008 to 2010, and a Principal Staff Research Engineer with the Multimedia Research Lab (MRL), Motorola Labs, from 2000 to 2008. He is currently a Senior Staff Researcher with the Core Networks Research, Huawei Technology USA, Bridgewater, NJ. His research interests include audio-visual analytics and machine learning with its application in large scale video repositories annotation, mining, and recommendation, as well as video adaptation, source-channel coding and distributed optimization issues of the wireless video networks. He has 12 issued or pending patents and 60+ publications in book chapters, journals, and conference proceedings in these areas.

Dr. Li was elected Vice Chair of the IEEE Multimedia Communication Technical Committee (MMTC) 2008–2010. He received the Best Poster Paper Award at IEEE International Conference on Multimedia & Expo (ICME), Toronto, ON, Canada, 2006, and theBest Paper (DoCoMo Labs Innovative Paper) Award at the IEEE IEEE International Conference on Image Processing (ICIP), San Antonio, TX, 2007.

Dacheng Tao (M'07) is Professor of computer science with the Centre for Quantum Computation and Information Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. He mainly applies statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance.

Prof. Tao has authored and coauthored more than 100 scientific articles in top venues including, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, NIPS, ICML, ICDM, AISTATS, IJCAI, AAAI, CVPR, ECCV, ACM T-KDD, and KDD, with the best theory/algorithm paper runner-up award at the IEEE ICDM'07.