Biologically Inspired Feature Manifold for Scene Classification

Dongjin Song and Dacheng Tao, *Member, IEEE*

***Abstract*—Biologically inspired feature (BIF) and its variations have been demonstrated to be effective and efficient for scene classification. It is unreasonable to measure the dissimilarity between two BIFs based on their Euclidean distance. This is because BIFs are extrinsically very high dimensional and intrinsically low dimensional, i.e., BIFs are sampled from a low-dimensional manifold and embedded in a high-dimensional space. Therefore, it is essential to find the intrinsic structure of a set of BIFs, obtain a suitable mapping to implement the dimensionality reduction, and measure the dissimilarity between two BIFs in the low-dimensional space based on their Euclidean distance. In this paper, we study the manifold constructed by a set of BIFs utilized for scene classification, form a new dimensionality reduction algorithm by preserving both the geometry of intra BIFs and the discriminative information inter BIFs termed Discriminative and Geometry Preserving Projections (DGPP), and construct a new framework for scene classification. In this framework, we represent an image based on a new BIF, which combines the intensity channel, the color channel, and the C1 unit of a color image; then we project the high-dimensional BIF to a low-dimensional space based on DGPP; and, finally, we conduct the classification based on the multiclass support vector machine (SVM). Thorough empirical studies based on the USC scene dataset demonstrate that the proposed framework improves the classification rates around 100% relatively and the training speed 60 times for different sites in comparing with previous gist proposed by Siagian and Itti in 2007.**

***Index Terms*—Biologically inspired feature, dimensionality reduction, image retrieval, manifold learning, scene classification.**

I. Introduction

SCENE classification is a key component for many practical applications, e.g., robotics path planning [43], video content analysis [8], content-based image retrieval [33], and video surveillance [1]. Usually, scene classification is very difficult because of the wide variety of scenes potentially to be classified. In particular, variations in lighting, different view angles, and dynamic backgrounds lead to obstacles in efficient learning and robust classification.

In recent years, a large number of approaches have been developed and we can classify them into four categories: low-level visual feature-based schemes, local feature-based schemes, local-global feature-based schemes, and biologically inspired feature-based schemes.

A. Low-Level Visual Feature-Based Schemes

Low-level visual feature-based schemes, which represent scenes by global visual information [2], e.g., color, texture, and shape, have been successfully utilized in indoor/outdoor, city/landscape, and forest/mountain applications. The color information is the most informative feature for scene classification because it performs better than texture and shape with respect to scaling, rotation, perspective, and occlusion [40]. Representative global color features are HSV color histogram [40] and color coherence vectors [30]. The texture information is another important cue for scene classification. Previous studies have shown that texture information according to the structure and orientation fits well with the model of human perception and so does the shape information. Representative texture features include the multiresolution simultaneous autoregressive models [26] and wavelet-based decompositions [6], [24]. One of the most popular shape features is the edge directional histogram [18]. Because global visual features are sensitive to small geometric and photometric distortions and they fail to work in spite of large changes in viewing conditions, occlusions, and clutters, local feature-based schemes are developed.

B. Local Feature-Based Schemes

Local feature-based schemes represent scene images with detected interest points (or regions) based on some descriptors. Usually, bag-of-features (or textons) is utilized to represent a scene image. For local features, there are two groups of components: detectors and descriptors. Detectors, e.g., Harris's corner detector [13], scale invariant feature transform (SIFT) [21], Harris-Laplace detector [27], and maximally stable extremal region (MSER) detector [25], look for points and regions to mimic human observers in locating elemental features in a scene image. To describe detected interest points or regions, descriptors, e.g., SIFT [21], steerable filters [11], Gabor functions [7], Varma-Zisserman model [44], and gradient location and orientation histogram [28], are utilized. Descriptors should be distinctive and invariant to geometric and photometric transformations. Because local feature-based schemes ignore the spatial information, local-global feature-based schemes are introduced.

C. Local-Global Feature-Based Schemes

Local-global feature-based schemes utilize both the global spatial information and the local descriptors of interest points (or regions) to represent scene images to achieve a robust

Manuscript received April 24, 2008; revised February 06, 2009. First published September 25, 2009; current version published December 16, 2009 This work was supported by Nanyang Technological University SUG (Project number M58020010). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hassan Foroosh.

The authors are with the School of Computer Engineering, The Nanyang Technological University, Singapore 639798 (e-mail: dctao@ntu.edu.sg).

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.2009.2032939

classification. For example, Lazebnik *et al.* [22] developed the spatial pyramid matching by partitioning an image into increasingly fine grids and computing histograms of local features inside each grid cell. Their experimental results show improvements over the bag of words model. Gökalp *et al.* [12] proposed bag-of-regions by considering both the "bag of individual regions" where each region is regarded separately and the "bag of region pairs" where regions are with particular spatial relationships. Experiments on the LabelMe dataset show the effectiveness of this method. Although these approaches have many advantages in scene classification, they are short of evidences from visual cognition and neuroscience. Recent research outputs show significances of biologically inspired features in image understanding.

D. Biologically Inspired Feature-based Schemes

Biologically inspired feature-based schemes classify scenes by mimicking the process of visual cortex in recognition tasks. Recent reports from both neuroscience and computer vision have demonstrated that biologically plausible features [16], [38] are attractive in visual recognition. For example, Serre *et al.* [41] proposed a biologically inspired model to mimic the human visual cortex for object recognition and achieved top performance in both object categorization and scene classification. Oliva *et al.* [29] presented a set of perceptual dimensions, termed Spatial Envelope, to represent the dominant spatial structure of a scene and demonstrated that the Spatial Envelope is effective in representing visual semantic categories. Siagian and Itti [37] developed a novel scene classification framework based on the gist model to mimic the cognition process associated with visual attention. In this model, intensity, color, and Gabor features are combined together for scene representation, both principal components analysis (PCA) and independent components analysis (ICA) are utilized for feature selection, and a three-layer neural network is used for classification.

Although Poggio and Bizzi [31] showed that C1 units correspond to complex cells in the visual cortex and they are effective for object recognition, C1 units ignore both the color and intensity information of an image. Although the gist feature [37] used for scene classification takes color, intensity, and orientation information into account, the orientation information extracted by Gabor filters do not fully correspond to complex cells in the visual cortex. In addition, these features are labeled samples drawn from a low-dimensional manifold and artificially embedded in a high-dimensional ambient space, so PCA and ICA utilized by Siagian and Itti [37] for dimensionality reduction is not a suitable choice. This is because PCA and ICA do not consider both the non-Euclidean property of biological features and the sample label information. It is worth emphasizing that PCA and ICA perform well for classification tasks when the class label information is not yet available.

In this paper, to effectively represent color scene images, we unify C1 units together with both the color and intensity information used in the scene classification scheme developed by Siagian and Itti [37]. To discover the intrinsic coordinate of the newly unified biologically inspired gist feature, we parameterize samples in the original high-dimensional ambient space based on the proposed *discriminative and geometry preserving projections* (DGPP). DGPP precisely models both the intraclass geometry and interclass discrimination and never meets the undersampled problem. Finally, to classify scenes, we utilize pairwise multiclass support vector machine (SVM) [5] because of its good generalization ability for classification. We term the proposed procedure for scene classification as the biologically inspired feature manifold (BIFM) framework. To justify the effectiveness of the proposed BIFM, we compare it with the scene classification algorithm proposed by Siagian and Itti [37] and show the improvement for classification accuracy is around 100%.

The rest of the paper is organized as follows. Section II introduces the proposed biologically inspired feature manifold (BIFM) framework, which describes the newly unified biologically inspired gist feature, details the novel discriminative and geometry preserving projections, and briefs the pairwise multiclass support vector machine for scene classification. In Section III, we brief the USC scene dataset for empirical study. In Section IV, we compare the proposed BIFM with the scene classification algorithm developed by Siagian and Itti, and Section V concludes.

II. Biologically Inspired Feature Manifold Framework

Gist is a human perception process which may happen in brain areas that are preferentially respond to "place" with restricted spatial layout [9]. The proposed BIFM framework simulates this process and can be divided into three main components (as shown in Fig. 1): the newly unified biologically inspired gist feature, the novel DGPP which preserves both the inter separation between different classes and within-class intra geometry characteristics, and the pairwise multiclass SVM classifier.

A. BIF: Newly Unified Biologically Inspired Feature

In this framework, we unify C1 units, color and intensity units as the biologically inspired feature for scene image representation. The C1 units are utilized here to replace the original orientation channel in the scene classification algorithm developed by Siagian and Itti [37]. This is because the orientation information extracted by Gabor filters does not fully correspond to complex cells in the visual cortex.

C1 Units (16 Feature Maps): The C1 units correspond to complex cells in the visual cortex [41]. By using a maximum operation, C1 units pool over S1 units and only keep the max response of a local area of S1 units from the same orientation and scale. The S1 units correspond to simple cells in S1 layer of the visual cortex. Gabor functions are similar to the receptive field profiles in the mammalian cortical simple cells so they are utilized for representing the feature in S1 layer. The Gabor mother function is $F(x,y) = \exp(-(x_0+\gamma^2 y_0^2)/(2\delta^2)) \times \cos(2\pi x_0/\lambda)$, wherein $x_0 = x\cos\theta + y\sin\theta, y_0 = -x\sin\theta + y\cos\theta$, the range of x and y decides the scales of Gabor filters, and θ controls orientations. Here we arrange the Gabor filters to form a pyramid with eight scales and span a range of sizes from 7×7 to 21×21 pixels with a step of two pixels. Four orientations are considered: 0°, 45°, 90°, and 135°. By convoluting the initial input image with these $8 \times 4 = 32$ Gabor filters, 32

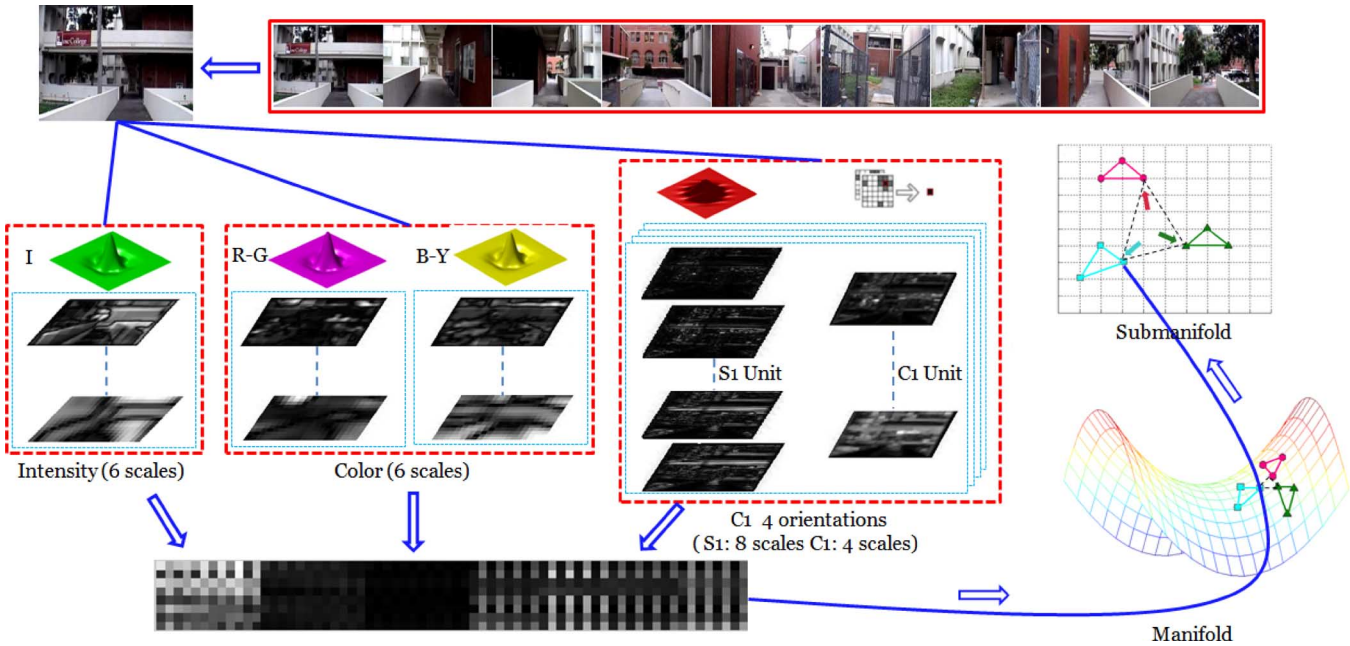

Fig. 1. Poposed biologically inspired feature manifold (BIFM) framework for scene classification: for each training scene image, we extract its biologically inspired features, which contain 34 feature maps, i.e., six intensity feature maps, 12 color feature maps, and 16 C1 units feature maps, and stack all extract 544 features as a vector. By using the proposed DGPP, we can project high-dimensional features to low-dimensional sub-manifold.

feature maps are obtained in S1 units to represent 32 different S1 receptive field types. Since the step between two consecutive scales of Gabor filters is small (two pixels), we can get precise orientation information on S1 feature maps.

These 16 feature maps of C1 units are obtained through pooling over two adjacent scales (with an identical orientation) of S1 units by using the maximum operation, i.e., a box (span a range of sizes from 8×8 to 14×14 with a step of two pixels) slides on the two adjacent scales of S1 units, and if a maximum intensity pixel is found the in the box, it will be used to represent the corresponding pixel in C1 feature map. Since the maximum operation shows some tolerance for shift and size and it also highlights the orientation information, it provides us precise and robust features.

Intensity Units (6 Feature Maps): The intensity units correspond to the neurons of mammals which are sensitive to dark centers on bright surrounds or bright centers on dark surrounds [17], [20]. By using dyadic Gaussian pyramids convolved on the intensity channel of an input color image, nine spatial scales are generated with a ratio from 1:1 (level 0) to 1:256 (level 8). To get intensity feature maps, the center-surround operation is performed between center levels ($c = 2, 3, 4$) and surround levels ($s = c + d$, with $d = 3, 4$, i.e., six feature maps are computed at levels of 2-5, 2-6, 3-6, 3-7, 4-7, and 4-8. Because scales are different between center levels and surround levels, maps of surround levels are interpolated to the same size as the corresponding center levels, and then they are subtracted point-by-point by the corresponding center levels to generate the relevant feature maps, i.e., $I(c, s) = |I(c) - \text{Interp}_{s-c} I(s)|$.

Color Units (12 Feature Maps): The color units are inspired by the "color double-opponent" system in cortex [17]. Neurons are excited by a color (e.g., blue) and inhibited by another color (e.g., yellow) in the center of receptive field, so are neurons in the surround. Herein, four color channels are used: $R = r - (g + b)/2, G = g - (r + b)/2, B = b - (r + g)/2$, and $Y = r + g - 2(|r - g| + b)$.

For each color channel (R, G, B, and Y), dyadic Gaussian pyramids are also used for generating nine spatial scales with a ratio from 1:1 (level 0) to 1:256 (level 8). The color pairs are formed as red-green (R-G) and blue-yellow (B-Y). The feature map, i.e., the across scales difference between two corresponding center and surround maps is also obtained by first interpolating the surround map to the same size of the relevant center map and then subtracted by the relevant center map point-by-point, i.e.,

$$\text{RG}(c, s) = |(R(c) - G(c)) - \text{Interp}_{s-c}(R(s) - G(s))| \text{ and}$$
$$\text{BY}(c, s) = |(B(c) - Y(c)) - \text{Interp}_{s-c}(B(s) - Y(s))|.$$

As shown by Siagian and Itti [36], color units are valuable for color image representation.

Scene Image Representation: In this paper, a color image is represented by 34 feature maps, wherein 16 feature maps obtained from C1 units, six intensity feature maps obtained from six intensity filters convoluted with the input image, and 12 color feature maps obtained from 12 color filters convoluted with the input image. We decompose each feature map into 4×4 grid sub-regions. All sub-regions have identical length and width. Then, the mean value of each sub region is calculated for final representation, as shown in Fig. 2, i.e., 16 mean values are utilized to represent each feature map and 544 values are obtained for image representation.

B. DGPP: Discriminative and Geometry Preserving Projections

In this framework, scene images are represented by the biologically inspired feature introduced in Section II-A and they

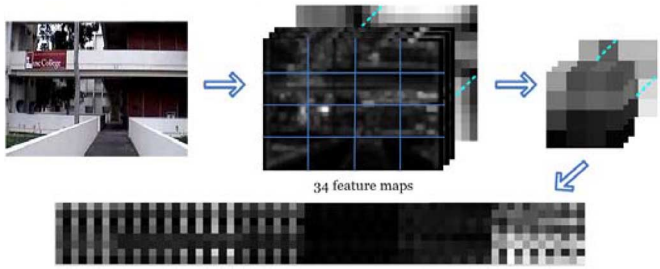

Fig. 2. Biologically inspired feature extraction procedure for a color scene image.

can be actually deemed as samples drawn from a low-dimensional manifold and artificially embedded in a high-dimensional ambient space. Here, the high-dimensional ambient space R^H is the biologically inspired feature space and the low-dimensional smooth manifold is R^L. Therefore, our objective is finding a mapping to select the effective subspace R^L from R^H for classification based on a number of observations $X = [\vec{x}_i]_{1\le i\le n} \in R^{H\times n}$. To reduce the complexity of the problem, we usually suppose the mapping is linear (i.e., the mapping is defined by a projection matrix $U \in R^{H\times L}$) and then we can find low-dimensional representations as $Y = U^T X \in R^{L\times n}$ and each column of Y is $\vec{y}_i = U^T\vec{x}_i \in R^L$.

In the past years, a number of manifold learning-based dimensionality reduction algorithms have been developed, including locally linear embedding (LLE) [32], ISOMAP [42], and Laplacian eigenmaps (LE) [3]. LLE uses linear coefficients, which reconstruct a given sample by its neighbors, to represent the local geometry, and then seeks a low-dimensional embedding, in which these coefficients are still suitable for reconstruction. ISOMAP, a variant of MDS, preserves global geodesic distances of all pairs of samples. LE preserves proximity relationships by manipulations on an undirected weighted graph, which indicates neighbor relations of pairwise samples. However, they suffer from the out of sample problem. One common response to address this problem is to apply a linearization procedure to construct explicit maps over new samples. Their corresponding linearization algorithms are neighborhood preserving embedding (NPE) [14], isometric projection, and locality preserving projections (LPP) [15].

However, for classification tasks, these algorithms only consider the intraclass geometry, while ignore the interactions of samples from different classes. Marginal Fisher analysis (MFA) [45] is a solution which takes both into account. However, it extracts discriminative information from only marginal samples, although nonmarginal samples also contain the discriminative information; and it preserves the local geometry of intraclass samples by making them as close as possible, i.e., the local geometry in MFA is described by the sample distribution compactness, but this is not a precise way for local geometry modeling. In addition, MFA has the undersampled problem when training samples are insufficient.

In this paper, we present a novel algorithm based on [48], which precisely models both the intraclass geometry and interclass discrimination and never meets the undersampled problem. That is in the low-dimensional space R^L, distances between interclass samples $\vec{y}_i$ should be as large as possible while distances between intraclass samples $\vec{y}_i$ should be as small as possible [39]; and the local geometry of intraclass samples should be preserved as much as possible by keeping linear reconstruction coefficients obtained in R^H [14], [19], [23], [32]. The formal part is termed the discrimination preservation and the latter part is termed the local geometry preservation, so we term the proposed algorithm as the discriminative geometry preserving projections (DGPP).

In DGPP, there are c classes, the ith sample $\vec{x}_i$ in the high-dimensional space is associated with a class label $m_i \in \{1, 2, \ldots, c\}$, the linear mapping is $U \in R^{H\times L}$, and U projects $\vec{x}_i \in R^H$ to a low-dimensional space as $\vec{y}_i = U^T\vec{x}_i \in R^L$. The lth class contains n_l samples, so $\sum_{l=1}^{c} n_l = n$.

To implement the discrimination preservation in DGPP, we would like to maximize the average weighed pairwise distance between samples in different classes and minimize the average weighed pairwise distance between samples in an identical class, i.e.,

$$\begin{aligned}\vec{y}_i &= \underset{\vec{y}_i, 1\le i\le n}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{j:m_j\neq m_i} \frac{s_{i,j}}{n}\|\vec{y}_i - \vec{y}_j\|^2 \\ &\quad - \sum_{i=1}^{n} \sum_{j:m_j=m_i} c_{i,j}\left(\frac{1}{n_{m_i}} - \frac{1}{n}\right)\|\vec{y}_i - \vec{y}_j\|^2 \\ &= \underset{\vec{y}_i, 1\le i\le n}{\operatorname{argmax}} \sum_{i,j=1}^{n} h_{i,j}\|\vec{y}_i - \vec{y}_j\|^2 \end{aligned} \tag{1}$$

where the weighing factor $h_{i,j}$ encodes both the distance weighing information and the class label information

$$h_{i,j} = \begin{cases} c_{i,j}(1/n - 1/n_l), & \text{if } m_i = m_j = l \\ s_{i,j}/n, & \text{if } m_i \neq m_j \end{cases} \tag{2}$$

where $c_{i,j}$ is set as $\exp(-\|x_i - x_j\|^2/\delta^2)$ according to LE and LPP for locality preservation and $s_{i,j}$ is set as 1 for simplicity. In addition, based on simple empirical justifications in this paper, $s_{i,j}$ affects the classification performance slightly, so it is reasonable to set it as 1.

To implement the local geometry preservation, we assume each sample can be reconstructed by the samples with the same label, i.e., $\vec{x}_i$ can be linearly reconstructed from $\vec{x}_j$ $(m_i = m_j)$ as

$$\vec{x}_i = \sum_{j:m_j=m_i} w_{i,j}\vec{x}_j + \vec{\varepsilon}_i \tag{3}$$

where $\vec{\varepsilon}_i$ is the reconstruction error for $\vec{x}_i$ and $w_{i,j}$ is obtained by minimizing $\sum_{i=1}^{n}\|\vec{\varepsilon}_i\|^2$, i.e.,

$$\begin{aligned} w_{i,j} &= \underset{w_{i,j}}{\operatorname{argmin}} \sum_{i=1}^{n} \|\vec{\varepsilon}_i\|^2 \\ &= \underset{w_{i,j}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \vec{x}_i - \sum_{j:m_j=m_i} w_{i,j}\vec{x}_j \right\|^2. \end{aligned} \tag{4}$$

By imposing $\sum_{j:m_j=m_i} w_{i,j} = 1$ and $w_{i,j} = 0$ (for $m_j \neq m_i$) on the above function, we have $w_i = \sum_p C_{i,p}^{-1} / \sum_{p,q} C_{p,q}^{-1}$,

wherein $C_{p,q} = (\vec{x}_i - \vec{x}_p)^T(\vec{x}_i - \vec{x}_q)$ is the local Gram matrix and $m_p = m_q = m_i$.

In DGPP, $w_{i,j}$ reconstructs $\vec{x}_i$ from $\vec{y}_j$ ($m_i = m_j$) in the low-dimensional space, so we have

$$\underset{\vec{y}_i}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \vec{y}_i - \sum_{j:m_j=m_i} w_{i,j}\vec{y}_j \right\|^2. \quad (5)$$

By combining the discrimination preservation and the local geometry preservation together, we have

$$\vec{y}_i = \underset{\vec{y}_i, 1\le i\le n}{\operatorname{argmax}} \sum_{i,j=1}^{n} h_{i,j}\|\vec{y}_i - \vec{y}_j\|^2 - \lambda \sum_{i=1}^{n} \left\| \vec{y}_i - \sum_{j:m_j=m_i} w_{i,j}\vec{y}_j \right\|^2. \quad (6)$$

Based on a series of matrix operations, we can obtain the linear projection matrix U according to

$$\begin{aligned} U &= \underset{U\in R^{H\times L}}{\operatorname{argmax}} \sum_{i,j=1}^{n} h_{i,j} \left\|U^T\vec{x}_i - U^T\vec{x}_j\right\|^2 \\ &\quad - \lambda \sum_{i=1}^{n} \|U^T\vec{x}_i - \sum_{j:m_j=m_i} w_{i,j}U^T\vec{x}_j\|^2 \\ &= \underset{U\in R^{H\times L}}{\operatorname{argmax}} \\ &\quad \times \left\{ \begin{array}{c} \operatorname{tr}\left[U^TX(D-H^T)(D-H^T)^TX^TU\right] \\ -\lambda\operatorname{tr}\left[U^TX(I-W^T)(I-W^T)^TX^TU\right] \end{array} \right\} \\ &= \underset{U\in R^{H\times L}}{\operatorname{argmax}} \operatorname{tr} \\ &\quad \times \left[U^TX \left(\begin{array}{c} (D-H^T)(D-H^T)^T \\ -\lambda(I-W^T)(I-W^T)^T \end{array} \right) X^TU \right] \\ &= \underset{U\in R^{H\times L}}{\operatorname{argmax}} \operatorname{tr}[U^TXMX^TU] \\ &= \arg \max_{U\in R^{H\times L}} \operatorname{tr}[U^T\tilde{M}U] \end{aligned} \quad (7)$$

where $H = [h_{i,j}] \in R^{n\times n}, W = [w_{i,j}] \in R^{n\times n}$, and $D \in R^{n\times n}$ is a diagonal matrix and its the ith entry is $\sum_{j=1}^{n} h_{i,j}$. The unified coefficient matrix $M = (D-H^T)(D-H^T)^T - \lambda(I-W^T)(I-W^T)^T$ is a symmetric matrix and $\tilde{M} = XMX^T$. By imposing $U^TU = I_d$ on (7), the solution is given by a standard eigenvalue decomposition on $\tilde{M}$ and U is formed by the L eigenvectors associated with the first L largest eigenvalues.

C. SVM: Support Vector Machine

SVM [5] is a popular pattern classification method used in recent years. It obtains top-level performance in different applications because of its good generalization ability in minimizing the Vapnik—Chervonenkis (VC) dimension and achieving a minimal structural risk.

We use the pairwise approach for SVM-based multiclass classification, so $C(C-1)/2$ classifiers are constructed and each classifier is trained by data from two different classes. In detail, for training data from the ith and the jth classes, we construct the following binary SVM classifier:

$$\begin{aligned} \underset{\vec{\alpha}\in R^m}{\operatorname{maximize}} W(\vec{\alpha}) &= \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j l_i l_j k(\vec{x}_i,\vec{x}_j) \\ \text{such that} \quad & 0 \le \alpha_i \le C, i = 1,\ldots,m \\ \text{and} \quad & \sum_{i=1}^{m}\alpha_i l_i = 0 \end{aligned} \quad (8)$$

where $\vec{x}_i \in R^L$ are training samples in the DGPP projected subspace, l is the class label ($+1$ for ith class and -1 for jth class) $\vec{\alpha}$ determines the classification hyper-plane to separate samples in ith class from samples in the jth class, $C > 0$ trades the complexity of the machine off the number of nonseparable samples, and m is the number of training samples from both the ith class and the jth class.

For a given input $\vec{x} \in R^L$, its label (i or j) is classified by $\operatorname{sgn}(\sum_{i=1}^{m} l_i\alpha_i k(\vec{x}_i,\vec{x}) + b)$, wherein the bias b is $1 - \sum_{i=1}^{m} l_i\alpha_i k(\vec{x}_i,\vec{x}_s)$ and $\vec{x}_s$ is a support vector with class label $+1$ (or it belongs to the ith class). During test, $C(C-1)/2$ times classification will be conducted and the voting rule is utilized for final decision: each binary classification can be deemed as a voting process where votes can be casted to $\vec{x}$ and $\vec{x}$ is designed to a class with maximum number of votes. In all experiments, the polynomial kernel is utilized.

III. USC Scene Dataset

All experiments for performance evaluation have been conducted on the University of Southern California (USC) scene dataset [37]. The USC scene dataset contains 375 video clips of three USC campus sites, which are Ahmanson Center for Biological Science (ACB), Associate and Founders Park (AnF), and Frederick D. Fagg Park (FDF). That is each site represents one scene. The ACB is a scene of buildings. Most of the video frames for ACB are flat walls with little texture and solid lines that describe walls and different parts of the buildings. The AnF is a park full of trees. It is dominated by vegetations and is full of complex texture. The FDF is an open field scene, which is composed by the sky, textureless space and random light. Each scene consists of nine segments, as shown in Fig. 3, with a number of video clips and each segment represents a path, a road or a hallway, i.e., different segments show different parts of a site. The task of the USC scene dataset is classification of video clip images over segments in each site.

In our experiments, we have two types of studies, i.e., small scale evaluation and large scale evaluation. Datasets for both small and large scales evaluation were obtained according to the following descriptions.

1) Dataset for small scale evaluation: for each site, we constructed four small datasets for four corresponding trials independently. Both training and testing sets were constructed by Siagian and Itti [37]. A small number of images are uniformly drawn from original training and testing video clips for training, and testing stages, respectively. The validation samples were also drawn from the original testing set but independent to the selected

Fig. 3. Example: Each image is selected from a segment in three USC sites. The rows from top to bottom are Ahmanson Center for Biological Science (ACB), Associate and Founders Park (AnF), and Frederick D. Fagg Park (FDF), respectively.

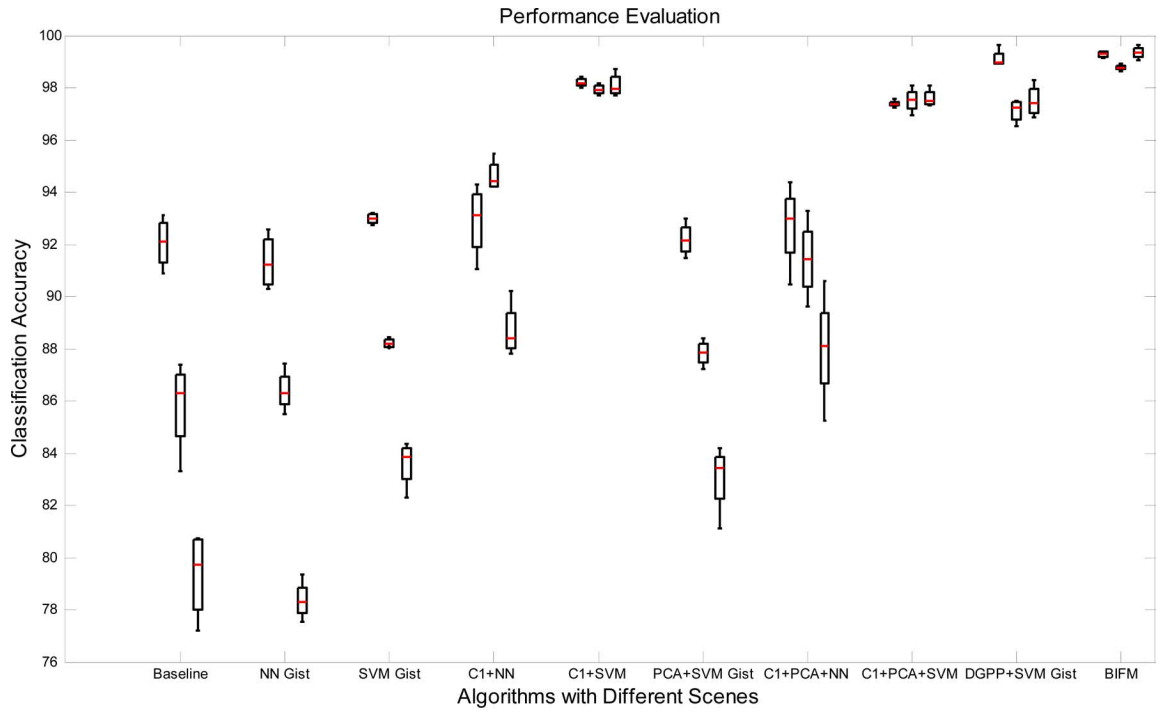

Fig. 4. Performance evaluation of different algorithms with different scenes.

testing samples. In ACB site, the cardinalities of training, validation, and testing sets for four trials are respectively: 1541/678/1356, 1284/678/1139, 1284/678/1136, and 1279/678/1133. In AnF site, the cardinalities of training, validation, and testing sets for four trials are respectively: 2827/1042/1041, 2825/1041/1038, 2820/1038/1037, and 2819/1037/1042. In FDF site, the cardinalities of training, validation, and testing sets for four trials are respectively: 1340/1467/1463, 1338/1463/1461, 1336/1461/1459, and 1335/1459/1467.

2) Dataset for large scale evaluation: the procedure for constructing both the training and testing video clips is identical to that used in Siagian and Itti's work [37]. The only difference is that in the test stage, they constructed four trails by dividing the video clips into four groups to test the performance of their system on different lighting conditions (i.e., late afternoon, early evening, noon, and mid-afternoon). We combined all test video clips of four lighting conditions together to test the performance of the proposed BIFM, and we uniformly sampled images from video clips with rate 5 frames per second. As a result, the cardinalities of training and testing for ACB site, AnF site, and FDF site are respectively 3847/3408, 11291/4158, and 10694/11694. For large scale evaluation, we did not use a validation set to choose parameters for DGPP because it performed robustly for different settings.

IV. Performance Evaluation

In this section, we evaluate the proposed BIFM in the following three aspects: 1) the effectiveness of each component in

BIFM for scene classification based on small scale evaluation datasets, 2) the effectiveness of the proposed DGPP with five representative subspace selection algorithms, e.g., PCA, linear discriminant analysis (LDA) [10], [46], NPE, MFA, and LPP, based on the ACB small scale evaluation dataset, and 3) the comparison between the proposed BIFM with Siagian and Itti's results based on the large scale evaluation datasets.

A. *Step-By-Step Model Justification*

This experiment justifies the effectiveness of each component in BIFM based on the standard image-based recognition procedure, i.e., feature extraction, subspace selection, and classification. This procedure is also utilized in Siagian and Itti's scene classification [37]. Therefore, the baseline utilized for comparison is Siagian and Itti's procedure. To justify the effectiveness of SVM for scene classification, we replace the three-layer neural network in the baseline with the SVM classifier and term this procedure as "PCA+SVM Gist". To justify the effectiveness of C1 units for scene classification, we replace the Gabor orientation channel in "PCA+SVM Gist" with C1 units and term this procedure as "C1+PCA+SVM". Similar to the procedure to justify the effectiveness of the C1 units, we justify the proposed DGPP in scene classification by replacing PCA/ICA stage in "PCA+SVM Gist" with DGPP and term this procedure as "DGPP+SVM Gist". We also show the performance of the baseline without the PCA step as "NN Gist", the performance of "PCA+SVM Gist" without PCA step as "SVM Gist", the performance of C1 units combined with NN as "C1+NN", the performance of "C1+PCA+SVM" without PCA as "C1+SVM", the performance of C1 units combined with PCA and NN as "C1+PCA+NN", and the performance of C1 units combined with PCA and SVM as "C1+PCA+SVM". Finally, we conduct the proposed BIFM for scene classification and compare it with all aforementioned procedures. In all experiments, both color and intensity features were used.

In all experiments, both DGPP and PCA/ICA reduced feature dimensionality from 544 to 80 according to suggestions from Siagian and Itti [37] for fair comparisons. The parameter λ in DGPP was set as 1 for all experiments because different settings affected the evaluation slightly. All procedures are conducted on three small scale evaluation datasets (four trials in each USC site). Experimental results are shown in Fig. 4.

Here, we utilize the boxplot to describe comparison results. Each boxplot produces a box and whisker plot for each method. The box has lines at the lower quartile, median, and upper quartile values. Whiskers extend from each end of the box to the adjacent values in the data-by default and the most extreme values within 1.5 times the interquartile range from the ends of the box. In Fig. 4, we have ten groups, each of which stands for a method, i.e., baseline developed by Siagian and Itti in [37], "NN Gist", "SVM Gist", "C1+NN", "C1+SVM", "PCA+SVM Gist", "C1+PCA+NN", "C1+PCA+SVM", "DGPP+SVM Gist", and BIMF. Each group contains three boxes, where boxes from left to right show the performances of ACB, AnF, and FDF, respectively.

As shown in Fig. 4, "SVM Gist", "C1 + SVM Gist", "DGPP + SVM Gist", and BIFM all perform better than the baseline developed by Siagian and Itti in [37]. Not only does the SVM Gist performs better than the baseline, it also has another two advantages in comparing with the baseline: 1) SVM provides a global optimal solution while the three-layer neural network utilized in the baseline can only achieve a local optimal solution, i.e., the baseline is sensitive to initial settings while SVM does not; and 2) the training cost of SVM is much less than that of the three-layer neural network, e.g., in this test, the

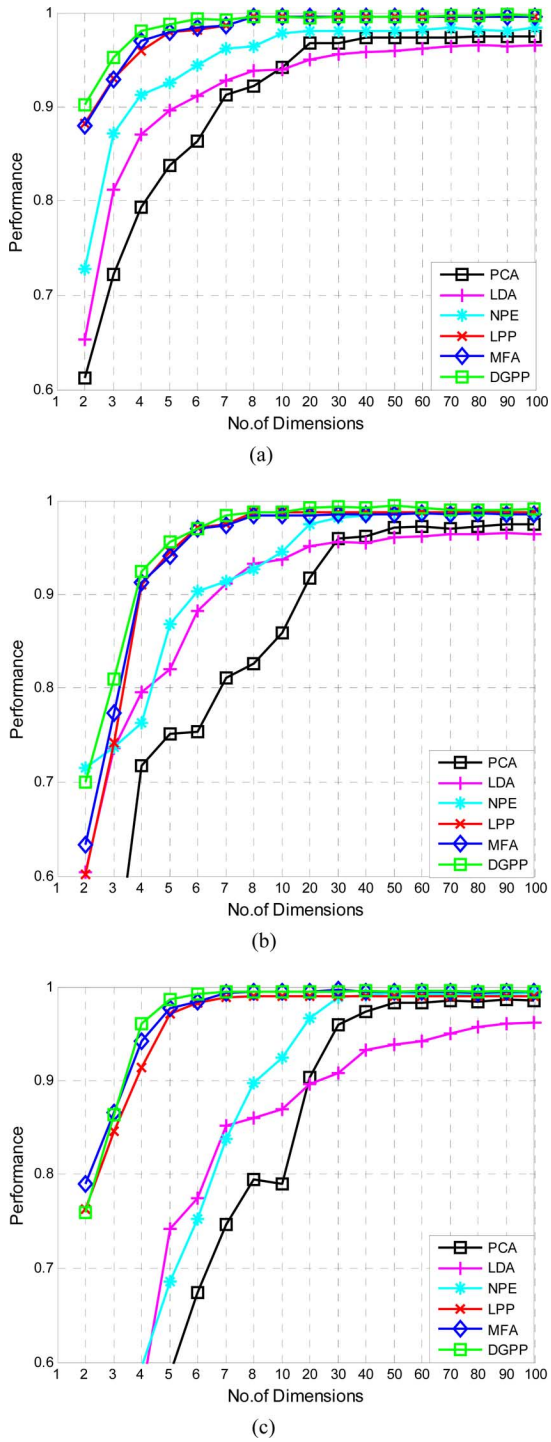

Fig. 5. Empirical studies of DGPP compared with five other representative subspace selection algorithms on ACB, AnF, and FDF. The y coordinate represents the classification accuracy and the x coordinate represents the dimensionality of selected subspace. (a) ACB, (b) AnF, (c) FDF.

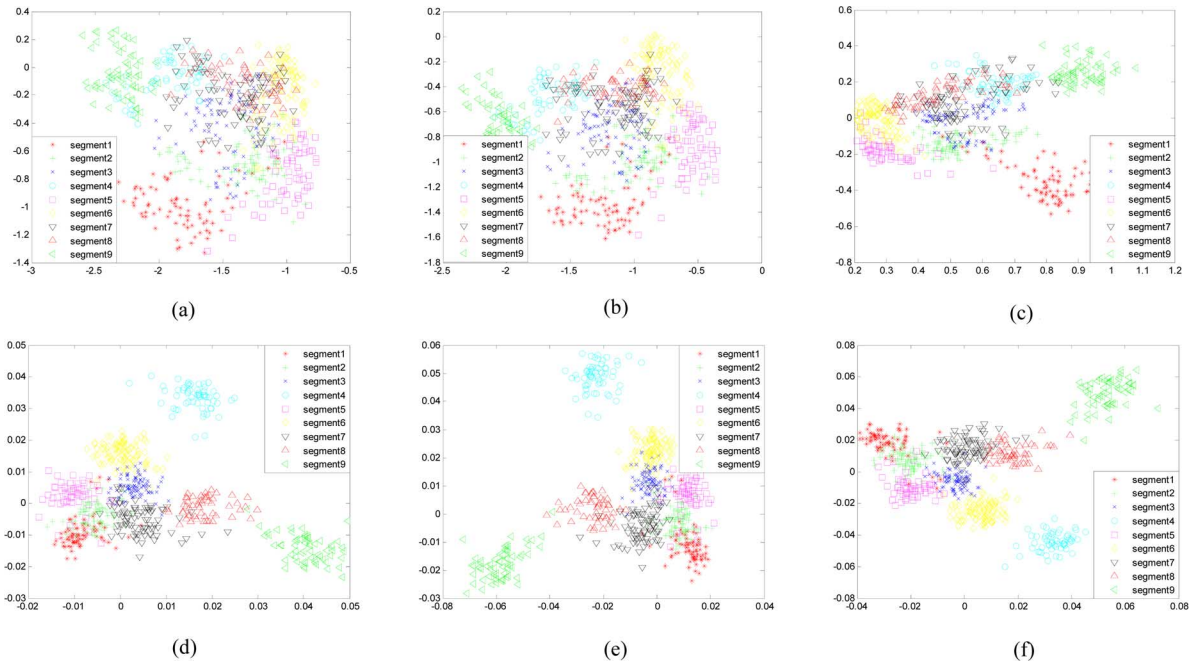

Fig. 6. Segment distributions in the selected two-dimensional subspace with different subspace selection algorithms on ACB. As shown in this figure, DGPP performs better than others. The classification accuracy (CA) is applied to measure different dimensionality reduction algorithms for separating classes in the projected subspace. (a) PCA ($\mathrm{CA} = 61.2\%$), (b) LDA ($\mathrm{CA} = 65.3\%$), (c) NPE ($\mathrm{CA} = 72.9\%$), (d) LPP ($\mathrm{CA} = 88.0\%$), (e) MFA ($\mathrm{CA} = 88.3\%$), (f) DGPP ($\mathrm{CA} = 90.3\%$).

training stage of the three-layer neural network is around one hour while the training stage of the SVM is around one minute.

C1 units can significantly improve the classification accuracy. C1 units pool over S1 units, which correspond to simple cells in the visual cortex, by using a maximum operation to keep only the max response of a local area of S1 units from the orientation and scale. The maximum operation is tolerant for affine transformations, so that C1 units provide more precise and robust feature than original orientation channel. Therefore, "C1 + SVM Gist" outperforms "SVM Gist".

DGPP performs much better than PCA/ICA in "SVM Gist". In this paper, scene images are represented by the biologically inspired feature, which can be actually deemed as samples drawn from a low-dimensional manifold and artificially embedded in a high-dimensional ambient space. Therefore, PCA/ICA is not suitable here for discovering the intrinsic structure because PCA/ICA only considers the Euclidean structure of samples. The proposed DGPP finds a mapping to transform the samples from the original high-dimensional ambient space (the biologically inspired feature space) to the low-dimensional smooth manifold so that it can discover the intrinsic structure of the samples and can achieve a significant improvement.

The combination of C1 and DGPP under "SVM Gist", i.e., BIFM, can achieve a further improvement in comparing with both of them and have an almost perfect performance (classification accuracy is more than 99%). Through our step-by-step empirical justification, we demonstrate that each of our three components, i.e., SVM classifier, C1 units, and DGPP for subspace selection, included in the proposed BIFM is effective in scene classification.

At the end of this experiment, it is worth emphasizing that "C1+SVM" is not as good as BIMF because DGPP can preserve the discriminative information and reduce sample noises. Although PCA can reduce sample noises, it fails to preserve the discriminative information, thus "C1+ PCA+SVM" is not as good as "C1+SVM".

B. DGPP Versus Representative Subspace Selection Methods

In this subsection, we justify the effectiveness of the proposed DGPP for subspace selection by comparing it with representative methods, e.g., PCA, LDA, NPE, LPP, and MFA, under the proposed BIFM framework for scene classification. It is worth noting that both LPP and NPE are utilized in supervised setting for fair comparison. According to our experiences, LPP and NPE with supervised setting perform better than them with unsupervised setting because unsupervised setting does not take the sample label information into account.

In all experiments, we justify feature dimensionality from 2 to 8 with step 1 and from 10 to 100 with step 10 in ACB, AnF, and FDF, respectively. All subspace selection algorithms are conducted on the small scale evaluation datasets (each with four trials) and the validation sets are applied for parameter selection for different algorithms. The classification accuracy (on testing sets) curves of different methods are shown in Fig. 5. As shown in these sub-figures, manifold learning algorithms, i.e., LPP, NPE, MFA, and DGPP, perform much better than conventional methods, e.g., PCA and LDA; and supervised learning algorithms, i.e., DGPP, MFA, LPP, NPE, and LDA, perform better than the unsupervised learning algorithm, i.e., PCA. Moreover, DGPP and MFA perform better than LPP and NPE because LPP and NPE for very low-dimensional projections (i.e., the dimensionality of selected subspace is less than or equals to 7) only consider the intraclass geometry while ignore the interactions of samples from different classes. Finally, DGPP usually performs

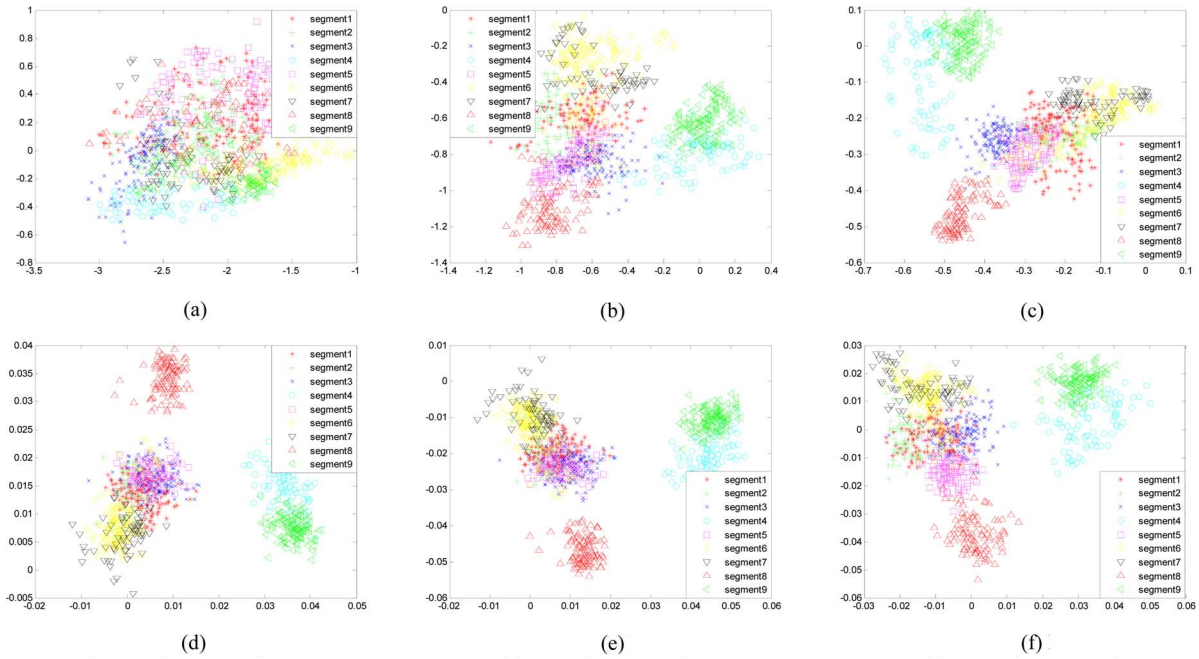

Fig. 7. Segment distributions in the selected two-dimensional subspace with different subspace selection algorithms on AnF. As shown in this figure, DGPP performs comparably to NPE and much better than others. The classification accuracy (CA) is utilized to measure different dimensionality reduction algorithms for separating classes in the projected subspace. (a) PCA ($\mathrm{CA} = 37.8\%$), (b) LDA ($\mathrm{CA} = 60.4\%$), (c) NPE ($\mathrm{CA} = 71.6\%$), (d) LPP ($\mathrm{CA} = 60.2\%$), (e) MFA ($\mathrm{CA} = 63.3\%$), (f) DGPP ($\mathrm{CA} = 70.3\%$).

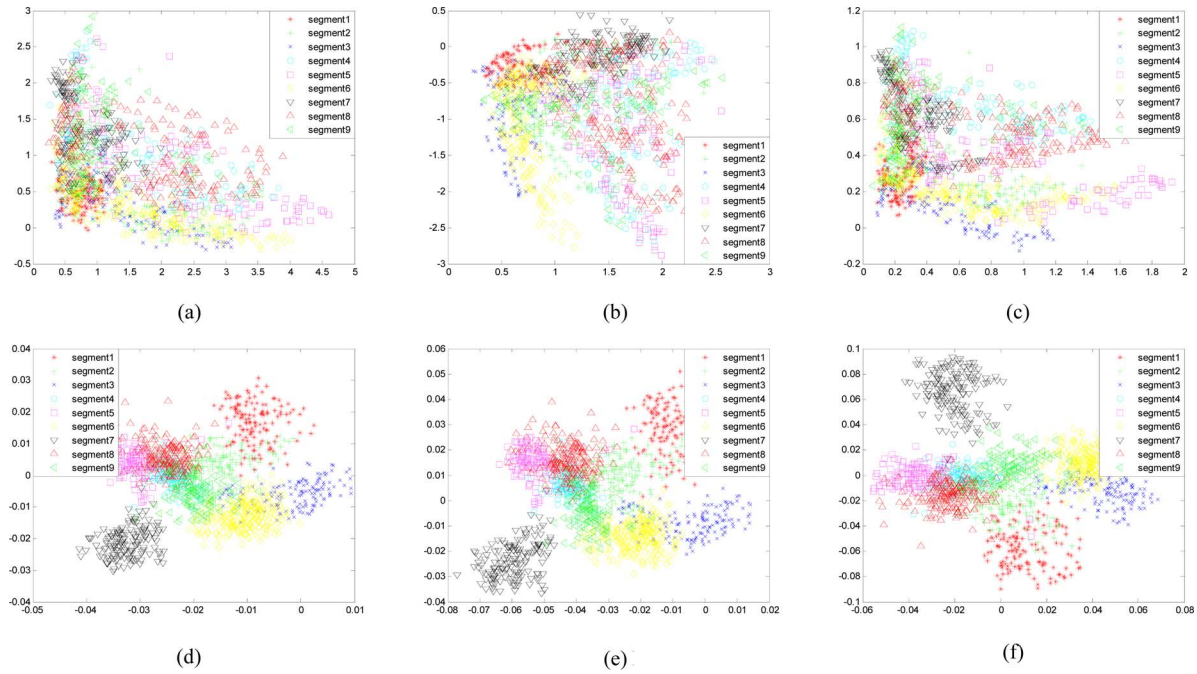

Fig. 8. Segment distributions in the selected two-dimensional subspace with different subspace selection algorithms on FDF. As shown in this figure, DGPP performs comparably to LPP, slightly worse than MFA, and much better than others. The classification accuracy (CA) is utilized to measure different dimensionality reduction algorithms for separating classes in the projected subspace. (a) PCA ($\mathrm{CA} = 24.2\%$) (b) LDA ($\mathrm{CA} = 28.6\%$) (c) NPE ($\mathrm{CA} = 29.0\%$), (d) LPP ($\mathrm{CA} = 76.3\%$), (e) MFA ($\mathrm{CA} = 79.0\%$), (f) DGPP ($\mathrm{CA} = 75.9\%$).

better than MFA for very low-dimensional projections (i.e., the dimensionality of selected subspace is less than or equals to 7). This is because 1) MFA extracts discriminative information from only marginal samples, although nonmarginal samples also contain the discriminative information; and 2) MFA preserves the local geometry of intraclass samples by making them as close as possible, i.e., the local geometry in MFA is described by the sample distribution compactness, whilst this is not a precise way for local geometry modeling.

Figs. 6–8 provide the subspace selection results, namely the sample distributions after they are projected onto the selected two-dimensional subspace on ACB, AnF, and FDF. The above points can be doubly justified by these figures. In detail, Fig. 6 shows that DGPP performs much better than other algorithms on

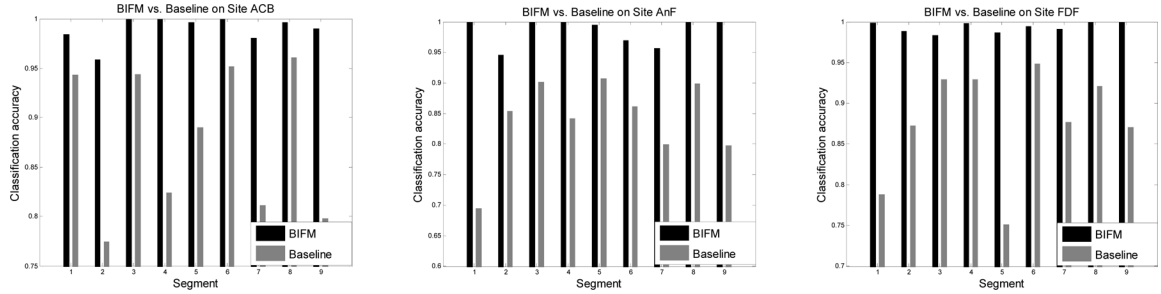

Fig. 9. Large-scale dataset-based performance evaluation. For each bar table, the y coordinate is the classification accuracy and x coordinate is the segment ID in a USC site. Gray histograms are obtained by the baseline and black histograms are obtained by the proposed BIFM.

ACB, and Figs. 7 and 8 show that DGPP performs comparably to other algorithms on AnF and FDF.

C. Performance Across Segments of All Sites

In this subsection, we compare the proposed BIFM with the baseline developed by Siagian and Itti in [37] based on the large scale evaluation dataset, which is described in Section III. Table I shows the classification results for all three sites in the USC dataset. Compared with the benchmark, the classification rate of each cite has been dramatically promoted up to around 99% (nearly 100%). We obtained a great improvement in classification accuracy in comparing with the baseline [37]. Furthermore, in the baseline, the classification results on AnF is much lower than the other two sites (about 3%–4%) due to the obstacles to classifier the vegetation dominated segments. However, BIFM can classify the vegetation dominated segments well. Fig. 9 illustrates the classification accuracy comparisons on each individual segment of the three sites based on BIFM and the baseline. In all experiments, we set λ as 1 because DGPP performed robustly for different settings in this experiment. Therefore, validation is not essential here.

V. Conclusion

In this paper, we proposed a novel biologically inspired feature manifold (BIFM) framework for scene classification, which contains three components: a new combination of popular biologically inspired features (BIFs) for scene image representation, a novel discriminative subspace selection method, termed discriminative geometry preserving projections (DGPP), for dimensionality reduction and a pairwise multiclass support vector machine (SVM) classifier for classification.

The new combination of BIFs mimics the human perception process in the visual cortex by considering C1, intensity and color units simultaneously. Because it is reasonable to assume BIFs are sampled from a low-dimensional manifold and embedded in a high-dimensional space, we proposed a novel manifold learning algorithm DGPP to map high-dimensional samples to a low-dimensional space. DGPP precisely preserves both the intraclass geometry and interclass discrimination and empirical studies show that it is superior to well known dimensionality reduction algorithms, e.g., principal component analysis (PCA), linear discriminant analysis (LDA), locality preserving projections (LPP), neighborhood preserving embedding (NPE), and marginal Fisher analysis (MFA), in the proposed BIFM for scene classification. Actually, the proposed DGPP can also be utilized for other applications, e.g., biometrics and multimedia information retrieval. The pairwise multiclass SVM is applied here for final classification because of its good generalization ability for classification. Thorough empirical studies based on the USC scene dataset showed the BIFM framework performs almost perfectly and significantly outperforms the current standard developed by Siagian and Itti [37]. In addition, BIFM is around 60 times faster than [37] in the training stage.

In the future, we would like to further improve our framework in the following aspects. First, we may bring new models for color and intensity information representation from visual cognitive science to make our BIFM more robust and can enable our system to classify more challenging scenes (based on the empirical studies in this paper, it is possible to enlarge the number of segments to at least 30). Second, we may extend the proposed framework to other applications, e.g., object categorization, detection and segmentation. Third, we will consider the unlabeled samples [47] for subspace learning and generalize DGPP to accept tensors [35] as input. Last but not the least, for visualizing samples in a 2-D space, we will consider using our recent geometric mean subspace method [36], [34] for biologically inspired feature analysis.

References

[1] B. Bose and E. Grimson, "Learning to use scene context for object classification in surveillance," presented at the IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2003.
[2] M. Boutell, C. Brown, and J. Luo, Review of the State of the Art in Semantic Scene Classification, Univ. Rochester. Rochester, NY, 2002.
[3] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 585–591, 2002.
[4] D. Cai, X. He, and J. Han, "Isometric projection," in *Proc. AAAI*, 2007, pp. 528–533.
[5] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines, 2001 [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm, See
[6] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Process.*, vol. 2, no. 4, pp. 429–441, Apr. 1993.
[7] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profile," *Vis. Res.*, vol. 20, pp. 847–856, 1980.
[8] N. Dimitrova, H.-J. Zhang, I. Sezan, T. S. Huang, and A. Zakhor, "Applications of video-content analysis and retrieval," *IEEE Multimedia*, vol. 9, no. 3, pp. 42–55, Mar. 2002.
[9] R. Epstein, A. Harris, D. Stanley, and N. Kanwisher, "The parahippocampal place area: Perception, encoding, or memory retrieval?," *Neuron*, vol. 23, pp. 115–125, 2000.
[10] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.
[11] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[12] D. Gökalp and S. Aksoy, "Scene classification using bag-of-regions representations," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2003, pp. 1–8.
[13] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vision Conf.*, 1988, pp. 147–151.
[14] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE Int. Conf. Computer Vision*, 2005, pp. 1208–1213.
[15] X. He and P. Niyogi, "Locality preserving projections," *Adv. Neural Inf. Process. Syst.*, vol. 16, 2004.
[16] Y. Huang, K. Huang, L. Wang, D. Tao, X. Li, and T. Tan, "Enhanced biologically inspired model," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
[17] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
[18] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognit.*, vol. 29, no. 8, pp. 1233–1244, 1996.
[19] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections," in *Proc. IEEE Int. Conf. Data Mining*, 2005, pp. 27–30.
[20] A. G. Leventhal, "The neural basis of visual function," in *Vision and Visual Dysfunction*. Boca Raton, FL: CRC, 1991, vol. 4.
[21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[22] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 2169–2178.
[23] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding. With high-order tensor data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 342–352, 2008.
[24] B. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
[25] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," in *Proc. British Machine Vision Conf.*, 2002, pp. 384–393.
[26] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognit.*, vol. 25, no. 2, pp. 173–188, 1992.
[27] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.
[28] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[29] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[30] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proc. ACM Int. Conf. Multimedia*, 1996, pp. 65–73.
[31] T. Poggio and E. Bizzi, "Generalization in vision and motor control," *Nature*, vol. 431, pp. 768–774, 2004.
[32] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
[33] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.
[34] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General averaged divergence analysis," in *Proc. IEEE Int. Conf. Data Mining*, 2007, pp. 302–311.
[35] 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.
[36] 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. 2008.
[37] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 300–312, Feb. 2007.
[38] D. Song and D. Tao, "C1 units for scene classification," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2008, pp. 1–4.
[39] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2005.
[40] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.
[41] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, Mar. 2007.
[42] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2 319–2 323, 2000.
[43] A. Ude and R. Dillmann, "Vision-based robot path planning," *Adv. Robot Kinematics Comput. Geometry*, pp. 505–512, 1994.
[44] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, no. 1–2, pp. 61–81, 2005.
[45] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[46] J. Ye, "Least squares linear discriminant analysis," in *Proc. Int. Conf. Machine Learning*, 2007, pp. 1087–1093.
[47] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proc. European Conf. Computer Vision*, 2008, pp. 725–738.
[48] T. Zhang, D. Tao, X. Li, and J. Yang, "A unifying framework for spectral analysis based dimensionality reduction," in *Proc. IEEE Int. Joint Conf. Neural Networks*, 2008, pp. 1671–1678.

Dongjin Song received the B.Eng. degree from the University of Science and Technology of China and the M.Phil. degree from the Hong Kong Polytechnic University.

He is currently a research assistant with the School of Computer Engineering, Nanyang Technological University. His research interests include computer vision, machine learning, and data mining.

Dacheng Tao (M'07) received the B.Eng. degree from the University of Science and Technology of China (USTC), the M.Phil. degree from the Chinese University of Hong Kong (CUHK), and the Ph.D. degree from the University of London, London, U.K.

Currently, he is a Nanyang Assistant Professor with the School of Computer Engineering, Nanyang Technological University, and holds a visiting post in London. He is a Visiting Professor at Xi Dian University and a Guest Professor at Wu Han University. His research is mainly on applying statistics and mathematics for data analysis problems in computer vision, multimedia, machine learning, data mining, and video surveillance. He has published more than 100 scientific papers in top venues with best paper runner up awards and finalists. He has authored/edited six books and eight journal special issues.

Dr. Tao is an associate editor of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING and *Computational Statistics & Data Analysis* (Elsevier). He has (co-)chaired special sessions, invited sessions, workshops, panels, and conferences.