Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models

Jia Li, *Member, IEEE*, and James Z. Wang, *Member, IEEE*

***Abstract*— This paper addresses learning based characterization of fine art painting styles. The research has the potential to provide a powerful tool to art historians for studying connections among artists or periods in the history of art. Depending on specific applications, paintings can be categorized in different ways. In this paper, we focus on comparing the painting styles of artists. To profile the style of an artist, a mixture of stochastic models is estimated using training images. The 2-D multiresolution hidden Markov model (MHMM) is used in the experiment. These models form an artist's distinct digital signature. For certain types of paintings, only strokes provide reliable information to distinguish artists. Chinese ink paintings are a prime example of the above phenomenon; they do not have colors or even tones. The 2-D MHMM analyzes relatively large regions in an image, which in turn makes it more likely to capture properties of the painting strokes. The mixtures of 2-D MHMMs established for artists can be further used to classify paintings and compare paintings or artists. We implemented and tested the system using high-resolution digital photographs of some of China's most renowned artists. Experiments have demonstrated good potential of our approach in automatic analysis of paintings. Our work can be applied to other domains.**

***Index Terms*— Image classification, image retrieval, art painting, mixture of stochastic models, 2-D multiresolution hidden Markov model**

I. Introduction

THERE have been many recent efforts to digitize fine art paintings and other art pieces. With the World-wide Web, it is now possible for anyone to gain access to digitized art pieces through the Internet. It is also becoming possible to analyze art works at a larger scale. Can we develop computer algorithms to analyze a large collection of paintings from different artists and to compare different painting styles? That is the question we attempt to address in this paper.

This is an important problem for not only computer scientists, but also for the art community. With advanced computing and image analysis techniques, it may be possible to use computers to study more paintings and in more details than a typical art historian could. Computers can be used to analyze fine features and structures in all locations of images efficiently. These numerical features can be used to compare paintings, painters, and even painting schools. Numerical features can also be used to assist database managers to classify and annotate large collections of images for effective retrieval purposes. The problem of studying paintings using computers is relatively new to the scientific community because paintings have been digitized at high resolutions since not long ago and advanced image analysis techniques are becoming available. Content-based image analysis and retrieval for stock photo databases has been studied extensively. However, the methods cannot be directly applied to the study of photographs of paintings. We discuss existing image retrieval approaches and major challenges later in this section.

In this paper, we present our approach to study collections of Chinese paintings. A mixture of 2-D multiresolution hidden Markov models (MHMMs) is developed and used to capture different styles in Chinese ink paintings. The models are then used to classify different artists. We conducted experiments using a database of high resolution photographs of paintings. The algorithms presented here can potentially be applied to digitized paintings of other cultures.

A. Major challenges

The majority of work on image analysis is based on realistic imaging modalities, including photographs of real world objects, remote sensing data, MRI scans, and X-ray images. A rough correspondence exists in these modalities between objects and regions of relatively homogeneous colors, intensities, or textures. These pictorial features extracted locally can be clustered in a vector space, yielding a segmentation of the image. The segmented regions form an effective abstraction of the image and can be compared efficiently across images. Many image retrieval systems [21], [26], with the core technical problem of measuring the similarity between images, rely on such vector clustering based segmentation. This approach is also taken in many systems for image classification [16] and detection of objects of interest [24]. The expressive nature of art work, however, breaks the link between local pictorial features and depicted objects. For instance, many Chinese paintings are in monochromic ink and sometimes do not even possess gradually changing tones.

Furthermore, art paintings demand unconventional image analysis tasks. For instance, a significant genre of ancient Chinese paintings are the so called "mountains-and-waters" paintings. This genre depicts mountains, trees (an integral part

Manuscript received January 15, 2003; revised July 10, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ioannis Pitas.

J. Li is with Department of Statistics, The Pennsylvania State University, University Park, PA 16802 USA (email: jiali@stat.psu.edu)

J. Z. Wang is with School of Information Sciences and Technology and Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802 USA. Corresponding author. (email: jwang@ist.psu.edu)

The Website http://wang.ist.psu.edu provides more information related to this work.

Digital Object Identifier

of mountains), rivers/lakes, and sometimes small pagodas and thatch cottages, as shown in Figures 10, 11. In terms of image content, there is little to compare among these paintings. An important aspect art historians often examine when studying and comparing paintings is the characteristic strokes used by artists [7]. Many impressionism masters formed their styles by special strokes [10]. These include the swirling strokes of Van Gogh and the dots of Seurat. Zhang Daqian[1], an artist of the late Qing Dynasty to modern China, is renowned for creating a way of painting mountains using broad-range bold ink wash.

It is of great interest to study how to mathematically characterize strokes, extract different stroke patterns or styles from paintings, and compare paintings and artists based on them. There is ample room for image analysis researchers to explore these topics. In this paper, we investigate the approach of mixture modeling with 2-D MHMMs [15]. The technique can be used to study paintings from different aspects. Our current experiments focus on profiling artists.

B. Our approach

We profile artists using mixtures of 2-D MHMMs. A collection of paintings by several artists is used to train mixtures of 2-D MHMMs. Every 2-D MHMM in a mixture model, referred to as a component of the mixture, is intended to characterize a certain type of stroke. Based on the trained models, methods can be developed to classify paintings by artist and to compare paintings and artists. The mixture of 2-D MHMMs is motivated by several reasons:

1) The 2-D MHMM characterizes statistical dependence among neighboring pixels at multiple resolutions. The spatial dependence among pixels is closely related to the stroke style. For instance, small dark strokes generate pixels that frequently transit between bright and dark intensities. Thin strokes tend to generate large wavelet coefficients at higher frequency bands than thick wash. Pixels in regions of diluted wash correlate more strongly across resolutions than those of well defined sharp strokes.
2) The multiresolution hierarchical representation of spatial dependence employed in 2-D MHMM enables computationally the analysis of relatively large regions in an image. This capability is important because patterns of strokes can hardly emerge in small regions. The computation advantage of 2-D MHMM over a single resolution 2-D HMM is discussed in [15].
3) The mixture of 2-D MHMMs trained for each artist can be used not only to classify artists but also to extract and characterize multiple kinds of stroke styles. Comparing with a pure artist classification system, the mixture model offers more flexibility in applications. For instance, a composition of an image in terms of stroke styles each specified by a 2-D MHMM can be computed. Paintings of a single artist can be compared based on the stroke composition.
4) A major difficulty in classifying images using generic classification methods such as decision trees is to define a set of features that efficiently represent an entire image. Much work has been done on extracting features for small regions in images. How to combine these local features to characterize a whole image is not obvious. Under the 2-D MHMM approach, the local features are summarized by a spatial model instead of an overall feature vector. The distributions of the local features and their spatial relations are embedded in the model. Experiments have been performed on classifying ten categories of photographs using 2-D MHMM, SVM (support vector machine) [22] with color histogram based features, and SVM with features extracted from segmented regions [6], [25]. The 2-D MHMM approach yields the highest classification accuracy for this application.

[1]Conventionally, in Chinese, the family name is placed first.

C. Related work

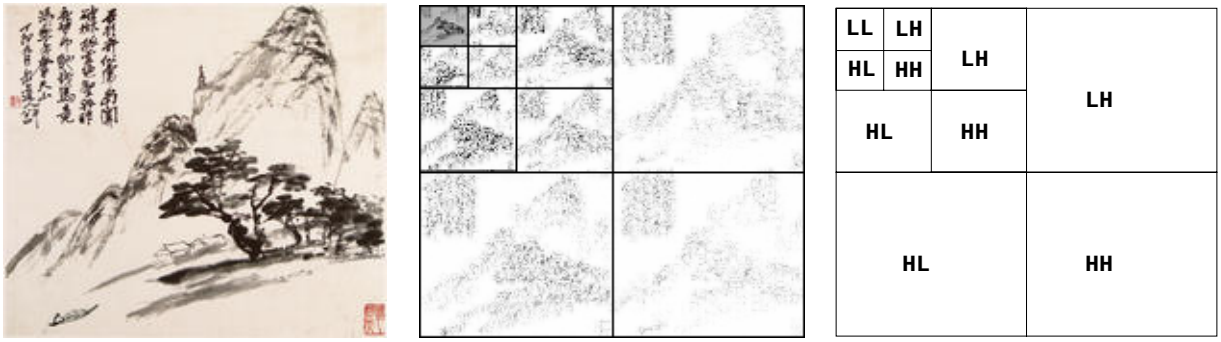
Research problems in concern and methodologies used in this paper are related to several technical fields, among which include computer vision, image retrieval, database management, and statistical image modeling. We do not intend a broad survey. Instead, we try to emphasize some work most related to what we propose. The references below should be taken as examples of related work, not as the complete list of work in the cited areas.

For a general introduction to digital imagery of cultural heritage materials and related database management, see [7], [18]. Readers are referred to [12], [26] for introduction to computer vision and image retrieval. Statistical image modeling has been explored extensively in both signal/image processing and computer vision. In 1991, Bouman and Liu used Markov random fields (MRF) for multiresolution segmentation of textured images [2]. Choi and Baraniuk [8] proposed wavelet-domain hidden Markov tree (HMT) models for image segmentation in 1999. Readers are referred to [2], [5], [8], [15] for an extensive review.

How to measure the similarity between images or groups of images is a core problem for content-based image retrieval (CBIR). An article published by Smeulders et al. reviewed more than 200 references [21] on the history of CBIR development. Readers are referred to that article and some additional references [1], [11], [13], [14], [19], [20], [26] for more information on recent advances in this area. CBIR systems have focused on comparing general color photographs. The techniques developed to measure similarity between such images are hardly suitable to the art images studied here.

D. Outline of the paper

The remainder of the paper is organized as follows. Section II describes how features are computed. In Section III, the mixture model of 2-D MHMMs is introduced and its estimation algorithm is presented. The architecture of the system for the particular application of classifying paintings of different artists is described in Section IV. In Section V, experiments and results are presented. Applications of the mixture of 2-D MHMMs other than classification are discussed in Section VI. We present our conclusions and suggest future research directions in Section VII.



Original image 3-level Wavelet transform Frequency bands

Fig. 1. The wavelet transform of an image. The wavelet transform decomposes an image into four frequency bands: LL, HL, LH, HH.

II. Feature Extraction

The wavelet transform is used to extract features from images. For a basic account on wavelet transforms, readers are referred to [9]. See [23] for applications in signal/image processing. By applying wavelet transforms successively to each row in an image and then each column, the image is decomposed into four frequency bands: LL (low low), HL (high low), LH, HH, as shown in Figure 1. The LL band contains a low resolution version of the original image. 2-D wavelet transforms can be applied recursively to the LL band to form multiresolution decomposition of the image. The HL band in large part reflects horizontal changes in the image. The LH band reflects vertical changes. The HH band reflects diagonal changes. Because of the dyadic subsampling, a block of size $\alpha \times \beta$ in the original image with top left corner located at coordinates (x, y) has its spatially corresponding blocks in the four bands with size $\frac{\alpha}{2} \times \frac{\beta}{2}$ and location $(\lfloor \frac{x}{2} \rfloor, \lfloor \frac{y}{2} \rfloor)$. The four frequency bands are usually spatially arranged in the manner shown in Figure 1 so that the transformed image is of the same dimension as the original image.

For the art images, features are extracted at three resolutions. The finest resolution (Resolution 3) is the original image. The coarsest resolution (Resolution 1) is provided by the LL band after two levels of wavelet transform. Daubechies 4 wavelet is used in particular because of its good localization properties and low computational complexity. Some other wavelet filters may generate similar results. The middle resolution is obtained by one level of wavelet transform. At each resolution, a feature vector is computed for every 2×2 block. Since the number of rows and that of columns decrease by a factor of two at a resolution one level coarser, the number of feature vectors reduces at a ratio of four across successively coarser resolutions. Figure 2 illustrates the pyramid structure of feature vectors extracted at the same spatial location in multiple resolutions. At a higher resolution, although the basic element of the image is still a 2×2 block, the spatial division of the image is finer because the image itself is expanded in both width and height. Equivalently, if we map images in all the resolutions to the original size, a 2×2 block in Resolution 1 corresponds to an 8×8 block in the original image. A feature vector extracted at Resolution 1 thus characterizes an original 8×8 block. At Resolution 2, this 8×8 block is divided into four child blocks of size 4×4, each characterized by a feature vector. These four child blocks are in turn divided at Resolution 3, each having its own four child blocks. To avoid terminology cumbersomeness, we refer to an 8×8 block in the original image and its corresponding 4×4 and 2×2 blocks in Resolution 2 and 1 as a pyramid. Each node in a pyramid denotes a 2×2 block at a certain resolution.

Next, we consider how to extract feature vectors at every resolution. Resolution 1 is used for description without loss of generality since the same mechanism of computing features is applied to all the resolutions. The feature vector for a 2×2 block includes the three wavelet coefficients at the same spatial location in the HL, LH, and HH frequency bands after one level of wavelet transform. Suppose the coordinates of pixels in the block are (x, y), $x = 2i, 2i + 1$, $y = 2j, 2j + 1$. The corresponding wavelet coefficients of this block in the LL, HL, LH, and HH bands are at coordinates (i, j), $(\frac{h}{2} + i, j)$, $(i, \frac{w}{2} + j)$, and $(\frac{h}{2} + i, \frac{w}{2} + j)$, where h and w are the numbers of rows and columns in the image. Since the low resolution images themselves are obtained by wavelet transforms, it is unnecessary to compute images in the three resolutions individually and then apply one-level wavelet transform to each one. Instead, we can simply apply a three-level wavelet transform to the original image and decompose it into an LL band and the HL, LH, and HH bands at the three resolutions, shown in the right panel of Figure 2. Wavelet coefficients in these frequency bands can be grouped properly to form the feature vectors at all the resolutions. Figure 2 shows how feature vectors are formed for the pyramid located at the top left corner. The three shaded pixels in the HL, LH, and HH bands at Resolution 1 denote the wavelet coefficients grouped into one feature vector for the node in the pyramid at the coarsest resolution. At the same spatial location at Resolution 2, there are four wavelet coefficients in every high frequency band. Consequently, 4 three-dimensional feature vectors are formed, each associated with one child node. Similarly, at Resolution 3, 16 three-dimensional feature vectors are formed, each associated with one node at the base of the pyramid.

For the current system implementation, as the focus is on Chinese ancient paintings, only the illuminance component of pixels is used. There are several reasons for discarding the color components. First, color was not considered as an essential element in painting by many traditional Chinese artists who basically used monochromic ink to paint. Painting

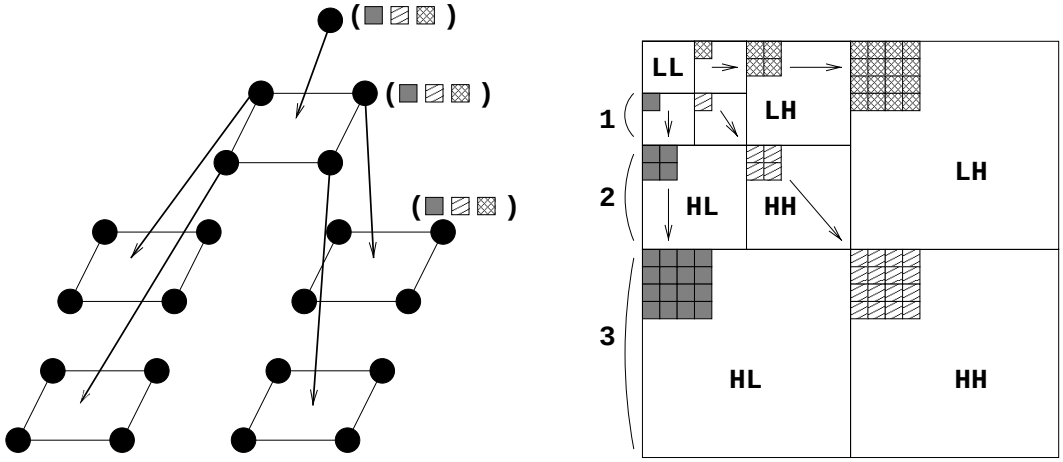


Fig. 2. The pyramid structure of the feature vectors across resolutions. Wavelet coefficients in high frequency bands are used as features.

was regarded as an integrated form of art with calligraphy, for which color is not used at all. Second, even when color was used in ancient Chinese paintings, there were rather limited varieties available, hardly enough to provide artists a sufficient amount of freedom to form their own distinct styles. Third, many paintings have serious color distortions caused by the aging over centuries. Various factors in the process of digitizing paintings add more color distortions. Finally, as pointed out by a reviewer of this paper, by not including color information in the features, we can study how well the algorithm works without color, which is interesting in its own right.

To reduce sensitivity to variations in digitization, as readers may have noticed, only high frequency wavelet coefficients, reflecting changes in pixel intensity rather than absolute intensity, are used as features. It is worth to point out, however, if color information is desired for characterizing images, it is straightforward to add in corresponding features. For instance, we can expand the feature vectors at the coarsest resolution to include the average color components of the 2×2 blocks. To incorporate color information, a three-level wavelet decomposition will be applied to each color component. The wavelet transform accounts for a majority part of computation in the feature extraction process. On a 1.7GHz Linux PC, the CPU time to convert a color image of size 512×512 to grayscale and compute the features described above using a three-level wavelet transform is about 0.83 second. The amount of computation is proportional to both the number of rows and the number of columns in an image.

III. Mixture of 2-D MHMM

A. Background on 2-D MHMMs

A detailed treatment of 2-D MHMMs can be found in [15]. The 2-D MHMM is proposed to capture the spatial dependence among image pixels or blocks and to explore the multiresolution nature of images. Under this model, an image is viewed as a 2-D stochastic process defined on a pyramid grid. Given a pixel representation of an image, multiple resolutions of the image are first computed. At every reduced resolution, the image size decreases by a factor of two in both rows and columns. A natural way to obtain a low resolution image is to use the LL (low low) frequency band yielded from a wavelet transform [9]. The one-level wavelet transform can be applied recursively to the LL band, giving representations of the image at successively coarser resolutions. A brief introduction to this process will be given in Section II.

Due to the localization property of the wavelet transforms, pixels in the image at multiple resolutions can be registered spatially to form a pyramid structure. To reduce computation, in the modeling process, the basic elements of an image may be non-overlapping blocks rather than pixels. Hence, terminologies to appear in the sequel are phrased in terms of blocks. The spatial registration of the blocks across resolutions and the pyramid abstraction are shown in Figure 3. A node in the pyramid at a certain resolution corresponds to a basic processing block of the image at that resolution. A block at a lower resolution covers a larger region of the image. As indicated by Figure 3, a block at a lower resolution is referred to as a parent block, and the four blocks at the same spatial location at the higher resolution are referred to as child blocks. We will always assume such a "quad-tree" split in this paper, since the extension to other hierarchical structures follows directly.

For every node in the pyramid, depending on particular applications, features may be computed based on pixel values in the node or in a neighborhood of the node. These features form a vector at the node and are treated as multivariate data. Details on feature extraction are presented in Section II. After applying feature extraction at all resolutions, the image is converted to a collection of feature vectors defined on a multiresolution pyramid grid. The 2-D MHMM attempts to model the statistical dependence among the feature vectors across and within resolutions.

The 2-D MHMM assumes that the feature vectors are probabilistic functions of an underlying state process defined on the pyramid. Given the state of a node, the feature vector is assumed to be conditionally independent of all other nodes in all resolutions. The conditional distribution of the feature vector is assumed to be multivariate Gaussian. The states are modeled by a multiresolution Markov mesh (a causal

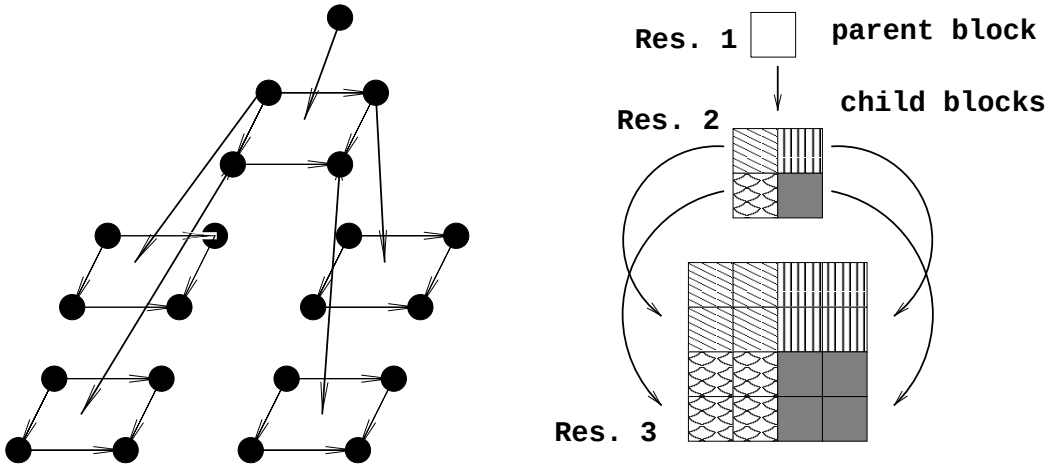


Fig. 3. The pyramid structure formed by spatial registration of blocks in multiple resolutions.

extension of Markov chain into two dimensions) [15]. They are purely conceptual and unobservable, playing a similar role as cluster identities in unsupervised clustering. In clustering analysis, samples are assumed independent, and hence so are the underlying cluster identities. For image analysis, since we intend to explore the spatial dependence, the states are modeled by a Markov mesh instead of an i.i.d. (independent and identically distributed) process, as normally assumed in clustering analysis. An important motivation for imposing statistical dependence among nodes through states, instead of directly on feature vectors, is to strike a good balance between model complexity and the flexibility of the marginal distribution of the feature vectors.

Next, we detail the assumptions made on the state process. First, let's consider a single resolution 2-D HMM. Denote the state at block (i,j) by $s_{i,j}$. We say that block (i',j') is before block (i,j) if either $i'<i$ or both $i'=i$ and $j'<j$, and write $(i',j')<(i,j)$. We assume that given the states of all the nodes before node (i,j), the transition probabilities of $s_{i,j}$ only depend on the states immediately above (i,j) and adjacent to the left of (i,j), i.e.,

$$P\{s_{i,j} \mid s_{i',j'} : (i',j')<(i,j)\} = P\{s_{i,j} \mid s_{i-1,j}, s_{i,j-1}\}\,.$$

For the multiresolution HMM, denote the set of resolutions by $\mathcal{R}=\{1,...,R\}$, with $r=R$ being the finest resolution. Let the collection of block indices at resolution r be

$$\mathbb{N}^{(r)} = \{(i,j):\, 0\le i<h/2^{R-r}, 0\le j<w/2^{R-r}\}\,,$$

where h or w is the number of blocks in a row or column at the finest resolution. An image is represented by feature vectors at all the resolutions, denoted by $u_{i,j}^{(r)}$, $r\in\mathcal{R}$, $(i,j)\in\mathbb{N}^{(r)}$. The underlying state of a feature vector is $s_{i,j}^{(r)}$. At each resolution r, the set of states is $\{1^{(r)},2^{(r)},...,M_r^{(r)}\}$. Note that as states vary across resolutions, different resolutions do not share states.

Statistical dependence across resolutions is assumed to be governed by a first-order Markov chain. That is, given the states at the parent resolution, the states at the current resolution are conditionally independent of the other preceding (ancestor) resolutions. The first-order dependence, in contrast to higher orders, is often assumed in multiresolution image models [8], [15] to maintain low computational complexity and stable estimation. By the chain rule of a Markov process, we have

$$\begin{aligned} & P\{s_{i,j}^{(r)} : r\in\mathcal{R}, (i,j)\in\mathbb{N}^{(r)}\} \\ = \;& P\{s_{i,j}^{(1)} : (i,j)\in\mathbb{N}^{(1)}\} \\ & \prod_{r=2}^{R} P\{s_{i,j}^{(r)} : (i,j)\in\mathbb{N}^{(r)} \mid s_{k,l}^{(r-1)} : (k,l)\in\mathbb{N}^{(r-1)}\}\,. \end{aligned}$$

At the coarsest resolution, $r=1$, states follow the Markov mesh assumed in a single resolution 2-D HMM. Given the states at resolution $r-1$, statistical dependence among blocks at the finer resolution r is constrained to sibling blocks (child blocks descended from the same parent block). Specifically, child blocks descended from different parent blocks are conditionally independent. In addition, given the state of a parent block, the states of its child blocks are independent of the states of their "uncle" blocks (non-parent blocks at the parent resolution). State transitions among sibling blocks are governed by Markov meshes, as assumed for a single resolution 2-D HMM. The state transition probabilities, however, depend on the state of their parent block. To formulate these assumptions, denote the child blocks at resolution r of block (k,l) at resolution $r-1$ by

$$\begin{aligned} & \mathbb{D}(k,l) = \\ & \{(2k,2l),(2k+1,2l),(2k,2l+1),(2k+1,2l+1)\}\,. \end{aligned}$$

According to the assumptions,

$$\begin{aligned} & P\{s_{i,j}^{(r)} : (i,j)\in\mathbb{N}^{(r)} \mid s_{k,l}^{(r-1)} : (k,l)\in\mathbb{N}^{(r-1)}\} \\ = & \prod_{(k,l)\in\mathbb{N}^{(r-1)}} P\{s_{i,j}^{(r)} : (i,j)\in\mathbb{D}(k,l) \mid s_{k,l}^{(r-1)}\}\,, \end{aligned}$$

where $P\{s_{i,j}^{(r)} : (i,j)\in\mathbb{D}(k,l)|\, s_{k,l}^{(r-1)}\}$ can be evaluated by transition probabilities conditioned on $s_{k,l}^{(r-1)}$, denoted by $a_{m,n,l}(s_{k,l}^{(r-1)})$. We thus have a different set of transition probabilities $a_{m,n,l}$ for every possible state in the parent

resolution. The influence of previous resolutions is exerted hierarchically through the probabilities of the states, which can be visualized in Figure 3.

As shown above, a 2-D MHMM captures both the inter-scale and intra-scale statistical dependence. The inter-scale dependence is modeled by the Markov chain over resolutions. Specifically, given the states of blocks at a parent resolution, the states of blocks at the child resolution are independent of resolutions preceding the parent. The intra-scale dependence is modeled by the HMM. At the coarsest resolution, feature vectors are assumed to be generated by a 2-D HMM. At all the higher resolutions, feature vectors of sibling blocks are also assumed to be generated by 2-D HMMs. The HMMs vary according to the states of parent blocks. Therefore, if the next coarser resolution has M states, then there are, correspondingly, M HMMs at the current resolution.

The 2-D MHMM can be estimated by the maximum likelihood criterion using the EM algorithm. The computational complexity of estimating the model depends on the number of states at each resolution and the size of the pyramid grid. Details about the estimation algorithm, the computation of the likelihood of an image given a 2-D MHMM, and computational complexity can be found in [15].

Since at the coarsest resolution, the states are related through a 2-D HMM, blocks in the entire image are statistically dependent. In practice, however, it is computationally expensive to assume a 2-D HMM over the whole image at Resolution 1 (the coarsest resolution). Instead, we usually divide an image into sub-images and constrain the HMM within sub-images. Sub-images themselves are assumed to be independent. For instance, an image contains 64×64 nodes at the coarsest resolution. Instead of assuming one HMM over the 64×64 grid, we may divide the image into $8 \times 8 = 64$ sub-images. At the coarsest resolution each sub-image contains 8×8 nodes modeled by an HMM. Consequently, the image is not viewed as one instance of a 2-D MHMM defined on the entire pyramid grid but as 64 independent instances of a 2-D MHMM defined on smaller pyramid grids. In fact, as long as the size of sub-images allows the analysis of sufficiently large regions, division into sub-images causes little adverse effect. Besides computation reduction, another advantage of using sub-images in contrast to treating the whole image will become clear in the next section where the mixture of 2-D MHMMs is introduced. Using the mixture of 2-D MHMMs, we can obtain a composition of an image in terms of different stochastic processes, a basis for comparisons between two images.

B. The Mixture Model

The purpose of using 2-D MHMMs to model art images is to capture the styles of artists' strokes. Capturing each style demands for the analysis of relatively large regions in the images. As it is constraining to assume that an artist has a single stroke style, we propose a mixture of 2-D MHMMs. For every sub-image, one of the component 2-D MHMMs is invoked and the feature vectors in the sub-image are assumed to follow the stochastic process specified by this MHMM. The idea parallels that of the Gaussian mixture [17]. When a single Gaussian distribution is insufficient to model a random vector, we may assume that the random vector is produced by multiple Gaussian sources. To produce the vector, a Gaussian source is randomly chosen. Then the vector is generated according to its distribution. Here, instead of being random vectors, every sub-image is a 2-D stochastic process. Therefore, every source specifies a 2-D MHMM rather than simply a Gaussian distribution. Based on the mixture model the 2-D MHMM most likely to generate a certain sub-image can be determined. Thus, a composition of an image in terms of the 2-D MHMMs can be obtained by associating each sub-image to the most likely mixture component. This composition is useful for detailed comparisons between images, a point to be elaborated upon in Section VI.

A mixture of K 2-D MHMMs, denoted by $\tilde{\mathcal{M}}$, is parameterized by the prior probabilities of the components, λ_k, and the individual 2-D MHMMs $\mathcal{M}_k$, $k = 1, 2, ..., K$. We denote the collection of feature vectors in a sub-image by $\mathbf{U}$. Then the probability of $\mathbf{U}$ under the mixture model is:

$$P(\mathbf{U} \mid \tilde{\mathcal{M}}) = \sum_{k=1}^{K} \lambda_k P(\mathbf{U} \mid \mathcal{M}_k) \ .$$

Assume the training sub-images are $\{\mathbf{U}_1, \mathbf{U}_2, ..., \mathbf{U}_n\}$. To estimate the mixture model, the standard EM procedure is used to update the $\mathcal{M}_k$'s iteratively by the following two steps:

1) E step: compute the posterior probabilities of the mixture components for each sub-image:

$$p_{k,i} \triangleq P(\mathcal{M}_k \mid \mathbf{U}_i) \ ,$$

where $P(\mathcal{M}_k \mid \mathbf{U}_i) = P(\mathcal{M}_k, \mathbf{U}_i)/P(\mathbf{U}_i) = \lambda'_k P(\mathbf{U}_i \mid \mathcal{M}_k) / \sum_{k'=1}^{K} \lambda'_{k'} P(\mathbf{U}_i \mid \mathcal{M}_{k'})$.

2) M step: update the prior probabilities

$$\lambda_k = \frac{\sum_{i=1}^{n} p_{k,i}}{n} \ ,$$

and the MHMMs $\mathcal{M}_k$, $k = 1, ..., K$ using weights $p_{k,i}$.

Note that λ'_k are prior probabilities estimated in previous iteration. The update of MHMMs with weights can be performed by the estimation algorithm for a single MHMM described in [15].

An alternative estimation to the maximum likelihood estimation by EM is the so-called classification maximum likelihood (CML) approach [17], which treats the mixture component identity of every sub-image as part of the estimation. The corresponding version of the EM algorithm is referred to as the classification EM [4] (CEM) algorithm. CEM modifies EM by replacing the "soft" classification into mixture components in the E step by a "hard" classification. After computing $p_{k,i}$, the ith sub-image is classified to the MHMM with maximum $p_{k,i}$ over k. In the M step, each MHMM is estimated using sub-images classified to it. Or equivalently, $p_{k,i}$ is set to 1 if $p_{k,i} > p_{k',i}$, for all $k' \neq k$ and 0 otherwise. CEM reduces the computation of EM by estimating each component model using only sub-images classified to this component. In contrast, EM estimates every component model based on all the sub-images, each weighted by a posterior probability.

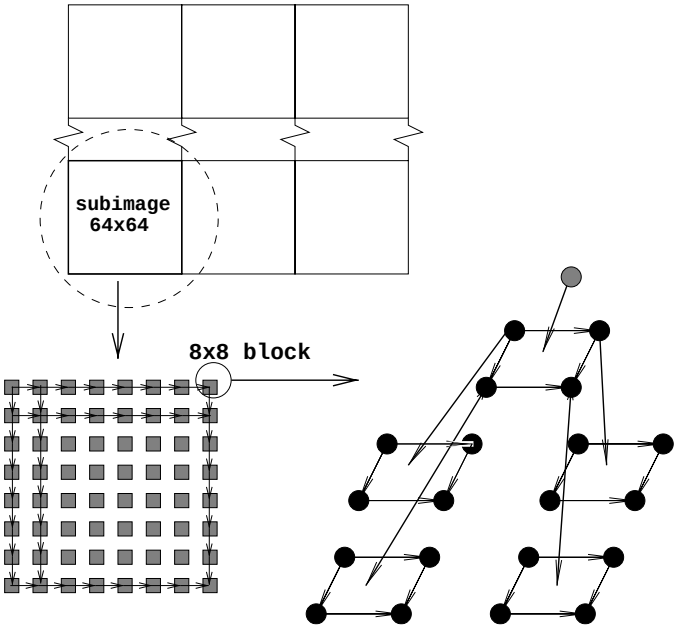


Fig. 4. The three resolution model used in the system. An image is divided into sub-images of size 64×64. For each 8×8 block in a sub-image, feature vectors extracted at the three resolutions form a pyramid shown above. The 8×8 blocks within the sub-image are statistically dependent because the root nodes of their pyramids are assumed to follow a 2-D HMM.

The advantage of CEM is especially appealing when it is computationally intensive to estimate the component models. The CEM algorithm is used in estimation in our current system.

IV. System Architecture

In this section, we introduce the architecture for classifying paintings of different artists. In training, a set of paintings from each artist is collected and multiresolution features are computed for each image. These multiresolution features, with their spatial location information, are input to estimate a mixture of 2-D MHMMs, profiling the artist. The spatial information is needed since the feature vectors are not treated as independent samples. On the other hand, the transition probabilities in the 2-D MHMM only depend on the relative spatial location of feature vectors between and within resolutions. The amount of computation needed to train the mixture model is proportional to the number of sub-images in the training images.

To classify an image, it is first converted to a set of multiresolution feature vectors. The classification of an image is based on the classification of its sub-images. The likelihood of a sub-image i, $i = 1, 2, ..., n$, under each component 2-D MHMM of each artist's profiling mixture model $\tilde{\mathcal{M}}_\gamma$, $\gamma = 1, ..., \Gamma$, is computed. The sub-image i is labeled by class γ_i if a component of $\tilde{\mathcal{M}}_{\gamma_i}$ yields the maximum likelihood among the components of all the mixture models. A majority voting scheme using the class labels of all the sub-images determines the class of the image.

Figure 4 illustrates the forming of the pyramid of feature vectors in our system. An image is divided into sub-images of size 64×64. Sub-images from the same artist are assumed to be generated by a fixed mixture model $\tilde{\mathcal{M}}$. However, the mixture component identity, i.e., which MHMM is active, varies with sub-images. Every 8×8 block in a sub-image becomes a 2×2 block at the coarsest resolution (total of 3 resolutions), which is the basic processing element associated with one feature vector. The 8×8 blocks are statistically dependent because their associated feature vectors at the coarsest resolution are governed by a 2-D HMM. At the coarsest resolution, the feature vector of an 8×8 block corresponds to a root node of the pyramid. At the next higher resolution, the root node splits into 4 child nodes, each in turn splitting to 4 nodes at the highest resolution. Hence, the basic processing elements from resolution 1 to 3 correspond to 8×8, 4×4, and 2×2 blocks in the sub-image. If more than three resolutions are modeled by the 2-D MHMM, sub-images of larger sizes can be analyzed. However, larger sub-images are not necessarily desirable since each sub-image is assumed to be generated by one component 2-D MHMM and possess a single stroke style.

V. Experiments

A. Background on the Artists

We developed the system to analyze artistic paintings. As initial experiments, we studied and compared Chinese artists' work. We digitized collections of paintings by some of the most renowned artists in Chinese history at spatial resolutions typically of 3000×2000 pixels. Figure 5 shows a random selection of six images from the database. To validate the proposed method, we first used collections of five artists. For each artist, about one third of his collected paintings in the database are used as training images to estimate the mixture model, and the rest are used as testing images to evaluate classification performance. The training and testing images are both scaled so that the shorter of the two dimensions has 512 pixels. As explained in Section IV, the basic processing element at the lowest resolution corresponds to a block of 8×8 pixels at the highest resolution. If the longer dimension of an image is not divisible by 8, a narrow band (the width is smaller than 8) of pixels at one side of the image is discarded to guarantee divisibility by 8.

A brief introduction of the artists is given below. Complying to the naming tradition of Chinese paintings, the following terminologies are used to refer to the main categories of Chinese paintings: mountains-and-waters (landscape), flowers (a.k.a. flowers-and-birds), trees-and-grass, human figures, and animals.

1) Shen Zhou (1427-1509) of the Ming Dynasty: There are 46 of his paintings in the database. Most of them are of the mountains-and-waters type; a small number of them are of flowers.
2) Dong Qichang (1555-1636) of the Ming Dynasty: There are 46 of his paintings in the database; all are of the mountains-and-waters type.
3) Gao Fenghan (1683-1748) of the Qing Dynasty: There are 47 paintings of his in the database: some of mountains-and-waters and some of flowers.
4) Wu Changshuo (1844-1927) of the late Qing Dynasty: There are 46 paintings of his in the database, all of flowers.

Fig. 5. A random selection of six images from the database.

5) Zhang Daqian (1899-1983) of the late Qing Dynasty to modern China: Zhang was one of the few artists of modern China who inherited comprehensive skills from mountains-and-waters painters in the Ming and Qing Dynasty. He painted diverse topics: mountains-and-waters, flowers (mainly lotus), and human figures. There are 91 paintings of his in the database, encompassing all main categories of Chinese paintings.

B. Extract Stroke/Wash Styles by the Mixture Model

For each of the five artists, a mixture model with 8 components is trained. Every component 2-D MHMM has 3 resolutions, with 2 states at the each resolution. As described in Section IV, images are divided into sub-images of size 64×64. Every sub-image is assumed to be generated by one component MHMM in the mixture. After a mixture model is trained for an artist, sub-images in his paintings can be classified to one of the component MHMMs by the criterion of maximum posteriori probability. Assuming equal priors on the component, the criterion is equivalent to choosing the MHMM that yields the maximum likelihood of a sub-image. Computation of the likelihood is discussed in Section III. As Zhang has relatively diverse painting styles and topics, we examine and compare in detail sub-images classified to the 8 MHMMs in his paintings. Figures 6, 7, and 8 show regions in his paintings classified to each of the 8 MHMMs. Since sub-images are the classification elements, all the regions extracted are groups of connected sub-images classified to the same MHMM. The 8 MHMMs appear to have captured relatively distinct stroke or wash styles, described in detail below.

1) Swift, thin strokes on relatively smooth background: This type appears mainly in paintings of lotus and human figures. Some regions with Chinese characters written in pale ink are also classified to this group.
2) Flat controlled strokes: These appear mostly in the paintings of mountains and waters.
3) Heavy and thick wash: This style is used mainly to paint lotus leaves and trees on mountains.
4) Straight, pale wash with some vertical lines: These are used mostly for painting rocks.
5) Smooth regions.
6) Small dark strokes: Regions with Chinese characters and trees painted in detail tend to be classified to this group.
7) Sharp lines and straight washes: This style is mainly used to paint rocks.
8) Pale and diluted wash: This is used mainly to convey a vague impression, such as mountains in the distance.

Next, we illustrate the numerical result of a trained mixture model by examining the first resolution of one component 2-D MHMM in the mixture. In particular, the mixture model with eight components trained on Zhang's paintings is used. The transition probabilities $a_{m,n,l}$, $m, n, l = 1, 2$, are listed in Table I. For this example, the state of a block tends to be the same as the state of its left neighbor, regardless of the state of the above neighbor. The tendency of staying in the same state as the left neighbor is stronger when the above neighbor is also in this state. The Gaussian distributions of the three features in both states are plotted in Figure 9. For all the three features, their variances in State 1 are significantly higher than those in State 2 respectively. The mean values, in contrast, are

all close to zero. This indicates that the states differ mainly in the energy, reflected by the squared values, in high frequency bands generated by the wavelet transformation.

TABLE I
THE TRANSITION PROBABILITIES BETWEEN THE TWO STATES AT THE FIRST RESOLUTION OF ONE COMPONENT 2-D MHMM.

Transition probabilities	left: 1 above: 1	left: 1 above: 2	left: 2 above: 1	left: 2 above: 2
state: 1	0.94	0.74	0.29	0.05
state: 2	0.06	0.26	0.71	0.95

C. Classification Results

Classification results obtained by mixture models with different numbers of components and a decision tree based method are compared. In particular, we tested the mixture of one 2-D MHMM, four 2-D MHMMs, and eight 2-D MHMMs respectively. For the decision tree method, every sub-image of size 64×64, instead of an entire image, is treated as a sample because the number of training images is very limited and it is difficult to extract an efficient feature vector for an entire image, especially one in grayscale. The training data set thus comprises feature vectors extracted for the sub-images in the training images. The class label of a feature vector is γ if it is computed from a sub-image in artist γ's paintings. To classify a test image, its sub-images are classified using the trained decision tree and a majority voting scheme is performed afterwards. CART [3] (Classification and Regression Trees) is used to train decision trees. Features for a sub-image are computed using the three-level wavelet decomposition shown in Figure 1. For each of the 9 high frequency bands (LH, HL, HH bands at the three levels), the average absolute value of the wavelet coefficients in this band is used as one feature.

Three cases of classification have been studied. The artists classified in case 1, 2, 3 are respectively: 1) Shen and Dong, 2) Shen, Dong, Gao, and Wu, 3) Shen, Dong, Gao, Wu, and Zhang. Shen and Dong are compared with each other because they were both artists in the Ming Dynasty who focused on mountains-and-waters painting. Zhang possessed diverse painting styles and topics. His paintings are most likely to be confused with the others' work, as will be shown shortly. We thus examined the classification results with and without him as one class.

Table II summarizes the classification accuracy for each artist using different methods. For all the three cases of classification, the highest average accuracy is achieved by profiling every artist using a mixture of 4 or 8 2-D MHMMs. Comparing with the decision tree method, the mixture modeling approach yields considerably better classification on average in the case of classifying Shen and Dong and the case of classifying the five artists. When a single 2-D MHMM is used to characterize each artist, the classification accuracy is substantially worse than that obtained by a mixture of 4 or 8 components in all the three cases. This reflects that a single 2-D MHMM is insufficient for capturing the stroke/wash styles of the artists.

To examine whether the features extracted using the three-level wavelet decomposition lead to better classification in comparison with those extracted using only the one-level decomposition, the decision tree method is also applied to sample vectors containing merely the features computed from the wavelet coefficients in the three high frequency bands formed by the one-level wavelet transform. The average classification accuracy achieved in the three cases is 72%, 59%, and 20% respectively. The three-level decomposition results in the higher accuracy of 77%, 68%, and 30%, as shown in Table II.

Table III provides the detailed classification result for the five artists obtained by the mixture model of 8 components. Each row lists the percentages of an artist's paintings classified to all the five artists. Numbers on the diagonal are the classification accuracy for each artist. The classification accuracy for Zhang is high (85%). However, other artists' paintings tend to be misclassified to his work, which is consistent with the fact he had diverse painting styles and topics. If classification is performed only among the other four artists, the accuracy for Wu, Gao, and Dong increases from 55% to 94%, 68% to 77%, and 52% to 65% respectively. That for Shen decreases slightly

TABLE II
COMPARING CLASSIFICATION RESULTS OBTAINED BY DIFFERENT METHODS IN THREE CASES. THE MIXTURE MODELS WITH 1, 4, AND 8 COMPONENT 2-D MHMMS ARE EXAMINED. EACH COLUMN LISTS THE CLASSIFICATION ACCURACY FOR EACH ARTIST OBTAINED BY A PARTICULAR METHOD.

Case 1				
Accuracy (%)	Eight components	Four components	One component	Decision tree
Shen	70	83	63	93
Dong	90	90	87	61
Average	80	**87**	75	77
Case 2				
Accuracy (%)	Eight components	Four components	One component	Decision tree
Shen	47	63	50	83
Dong	65	58	68	32
Gao	77	65	16	58
Wu	94	97	97	97
Average	**71**	**71**	58	68
Case 3				
Accuracy (%)	Eight components	Four components	One component	Decision tree
Shen	50	70	53	13
Dong	52	52	65	0
Gao	68	61	16	19
Wu	55	97	97	23
Zhang	85	25	0	97
Average	**62**	61	46	30

TABLE III
THE CLASSIFICATION RESULT OBTAINED BY THE MIXTURE OF 8 2-D MHMMS. EACH ROW LISTS THE PERCENTAGES OF AN ARTIST'S PAINTINGS CLASSIFIED TO ALL THE FIVE ARTISTS. NUMBERS ON THE DIAGONAL ARE THE CLASSIFICATION ACCURACY FOR EVERY ARTIST.

Percent (%)	Shen	Dong	Gao	Wu	Zhang
Shen	50	17	30	0	3
Dong	0	52	13	0	35
Gao	6	13	68	0	13
Wu	0	0	3	55	42
Zhang	8	0	2	5	85

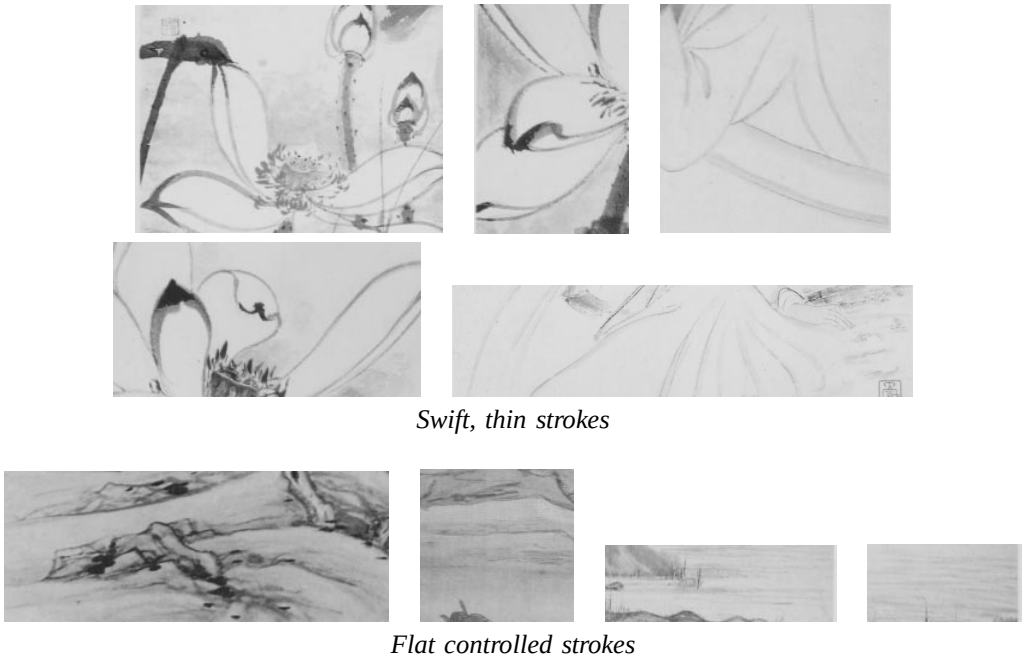


Fig. 6. Regions in Zhang's paintings classified as stroke style $1 \sim 2$.

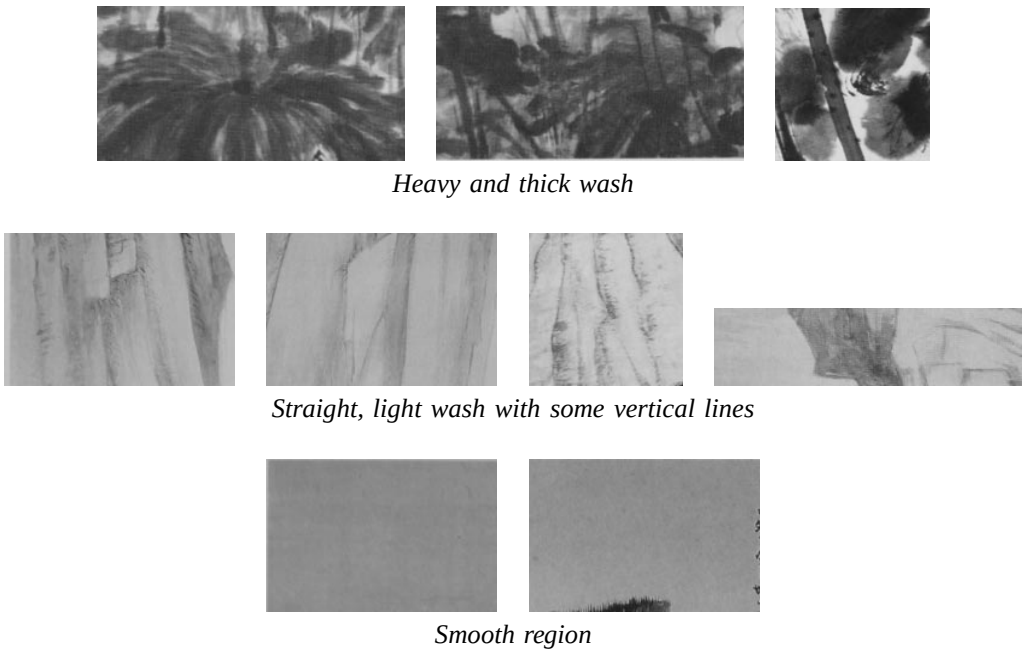


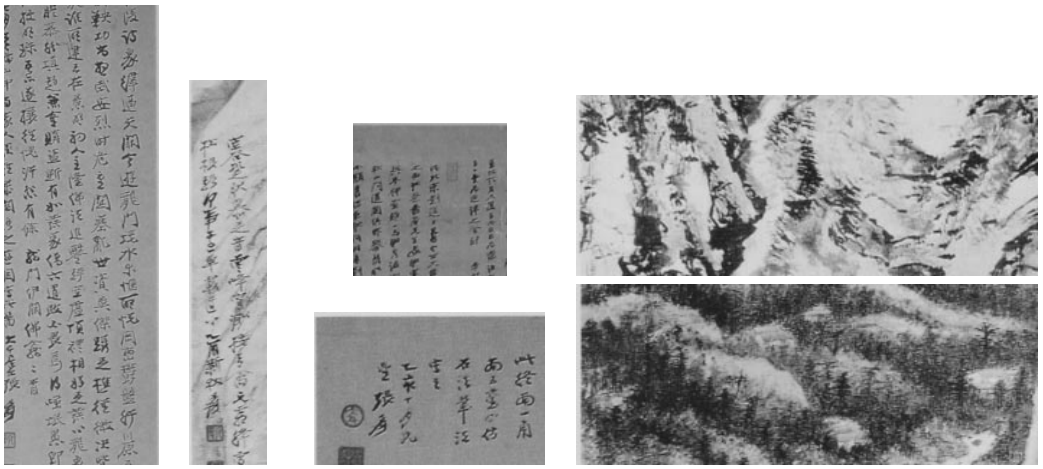
Fig. 7. Regions in Zhang's paintings classified as stroke style $3 \sim 5$.

from 50% to 47%. Zhang's diverse stroke styles require a relatively large number of mixture components to capture. Table II shows that if only four components are used, most images of Zhang are misclassified to the other artists.

VI. Other Applications

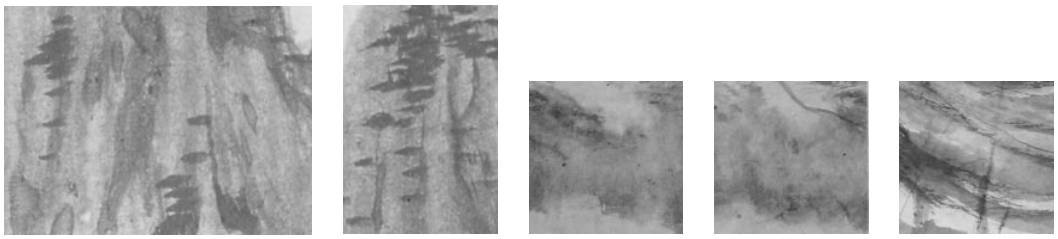
In the previous section, the mixture of 2-D MHMMs is used to classify images into their artists. With a set of trained mixture models, we can perform other types of analysis. A few possibilities are discussed in this section.

Comparisons among paintings can be made within one artist's work. Suppose we index the MHMMs in each mixture model from 1 to 8 (recall that 8 is the number of MHMMs in a mixture used in the experiment). By classifying every sub-image in an image into one of the MHMM in the mixture model, a map of the image into an array of MHMM indices (referred to as the stochastic model map) is obtained. These stochastic model maps form a basis for comparing images. A basic approach is to examine the percentages of sub-images classified to each MHMM. Consider two images $i = 1, 2$. Suppose the percentage of sub-images in image i classified



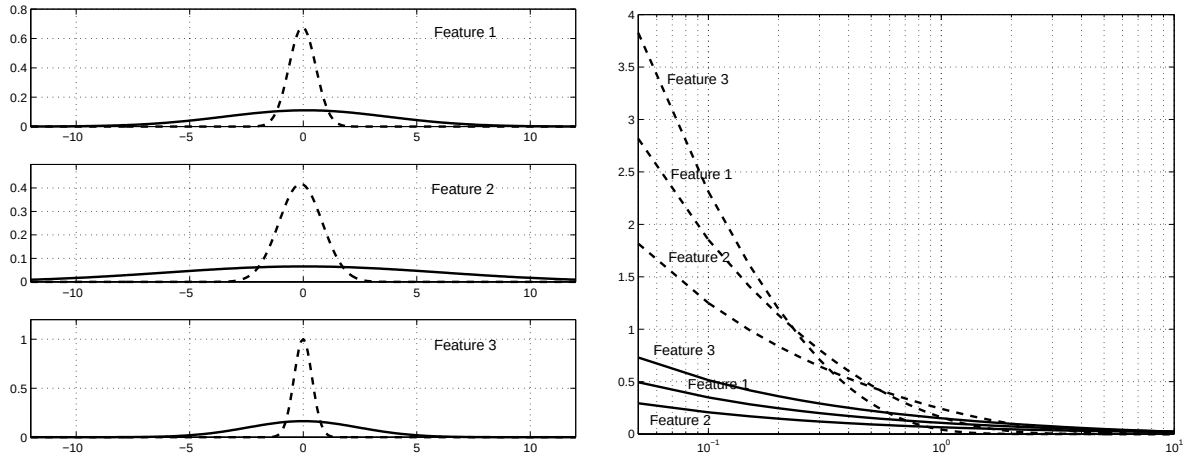
Small dark strokes

Sharp lines and straight wash



Pale and diluted wash

Fig. 8. Regions in Zhang's paintings classified as stroke style $6 \sim 8$.



(a) (b)

Fig. 9. The distributions of the three features in the two states at the first resolution of one component 2-D MHMM. (a): In every panel, the Gaussian distributions of one feature in both states are plotted. Solid lines: State 1, dash lines: State 2. (b): By approximating the means of the Gaussian distributions by zero, the square of each feature follows a scaled χ^2 distribution. The distributions of the squares of the three features in both states are plotted.

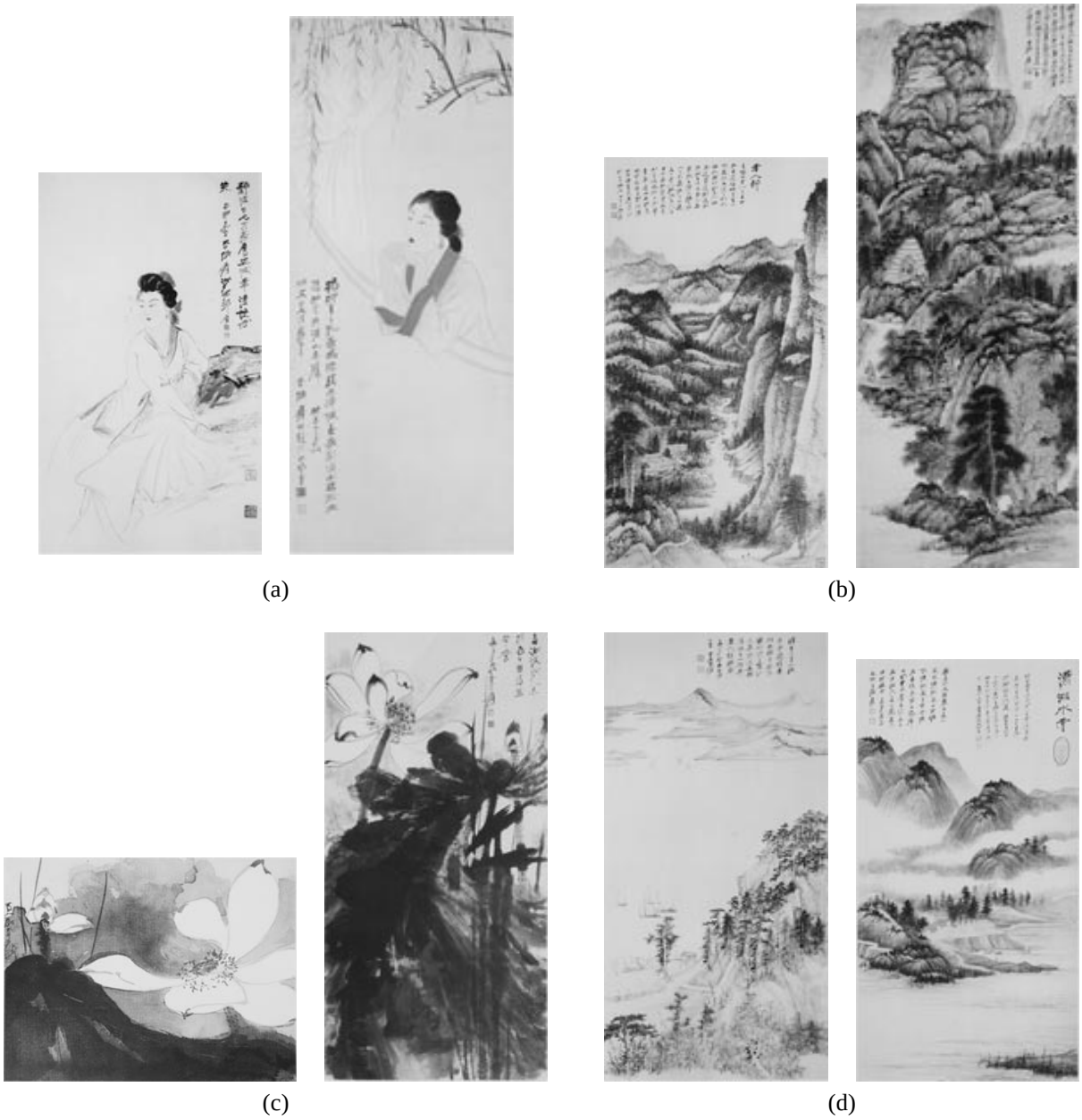


Fig. 10. Four pairs of Zhang's paintings. For each pair, the right image is the most similar one to the left according to the J "distance".

as the jth MHMM in the mixture is $p_{i,j}$, $\sum_{j=1}^{8} p_{i,j} = 1$. Similarity between images can be measured according to the closeness of the two probability mass functions (pmfs) $p_{1,j}$ and $p_{2,j}$, $j = 1, ..., 8$. A well-known measure of the disparity between two pmfs is the relative entropy defined as

$$I(p_1, p_2) = \sum_j p_{1,j} \log \frac{p_{1,j}}{p_{2,j}} .$$

$I(p_1, p_2)$ is nonnegative and equals zero if and only if p_1 and p_2 are identical. Another "distance" tested is formulated as follows:

$$J(p_1, p_2) = 1 - \sum_j \min(p_{1,j}, p_{2,j}) .$$

The "distance" $J(p_1, p_2)$ ranges from 0 to 1 and equals 0 if and only if p_1 and p_2 are identical and 1 if they are orthogonal. In Figure 10, 4 pairs of Zhang's paintings are shown. For each pair, the right image is the most similar one to the left according to the J "distance". The relative entropy yields the same result except for the image on the top left, for which another human figure image is chosen as the most similar one.

To obtain a crude numerical assessment of the similarity measures, we divide Zhang's paintings into three categories: mountains-and-waters, human figure, and lotus flowers. For each image, the most similar image is chosen from the rest according to the relative entropy or the J "distance". If the two images are in the same category, we mark a match. By deciding whether a match is achieved for every image of the artist, the total number of matches can be compared with the expected number of matches if random drawing is

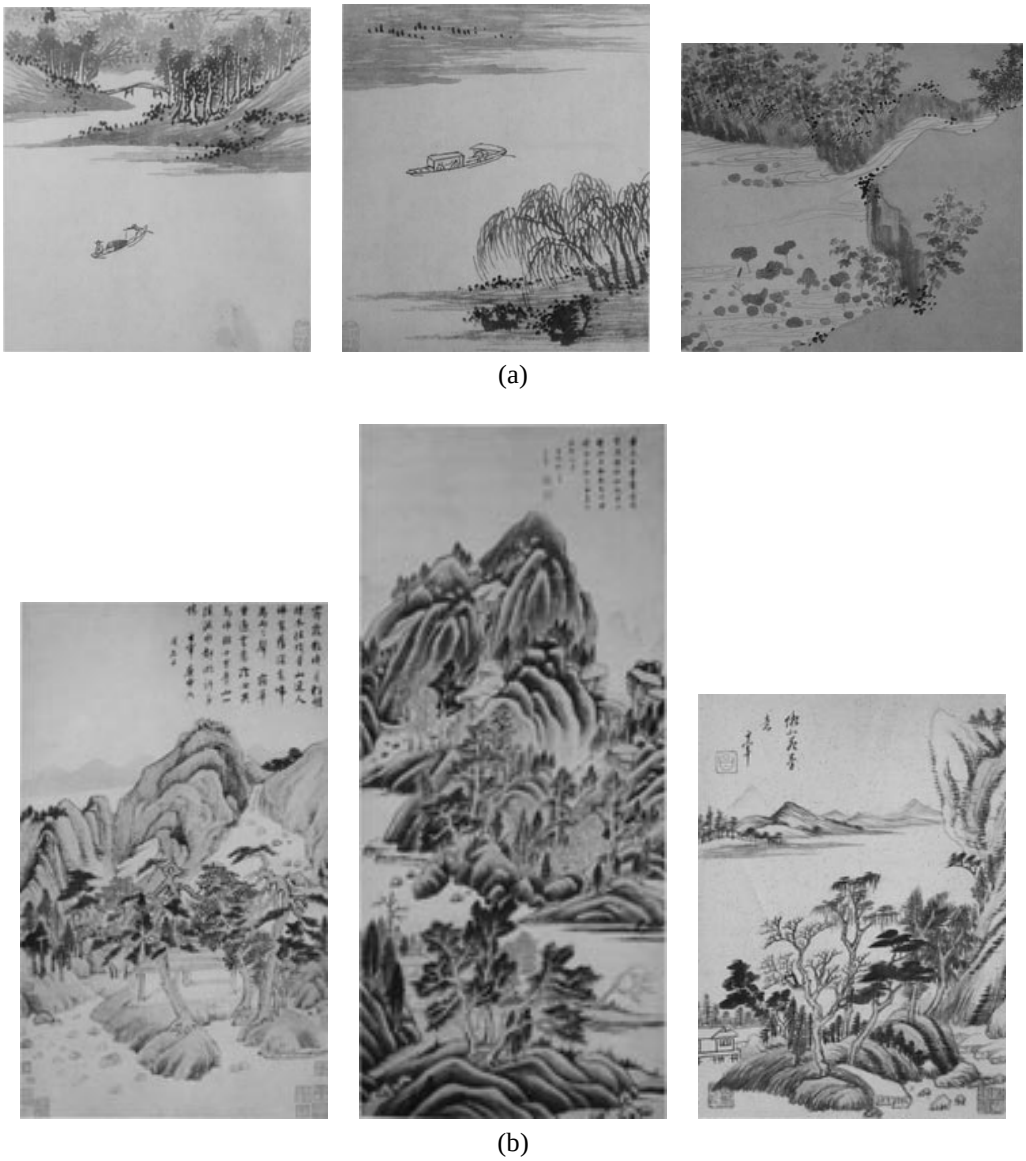


Fig. 11. Comparison between Shen and Dong's paintings. (a)/(b): Paintings of Shen/Dong identified as most similar to the other's work.

used. The number of images in each of the three categories, mountains-and-waters, figure, flowers, is $n_1 = 54$, $n_2 = 21$, $n_3 = 16$ respectively (total of $n = 91$). The expected number of matches is 38.7, computed based on the following formula:

$$\sum_{j=1}^{3} \frac{n_j(n_j - 1)}{n} \quad .$$

The number of matches provided by the relative entropy "distance" is 45; that by the J "distance" is 49.

When we compare two artists, it is often desirable to find paintings of each that most (or least) resemble those of the other. Since the artists are profiled by mixture models, to compare a painting of one artist versus the overall style of the other, we compute the likelihood of the painting under the profiling mixture model of each artist. Close values of the two likelihoods, or even a higher likelihood given by the other artist's model, indicate that the painting may be similar to the work of the other artist.

Figure 11 shows three mountains-and-waters paintings from Shen as well as Dong. The three from Shen are identified by computer as most similar to the style of Dong, and vice versa. Under the same spirit, we can find paintings of an artist most distinguished from the other artist's work. Figure 12 shows paintings of Zhang that are most different from Shen's work. Human figures and lotus flowers are painted in these images, which are indeed topics only depicted by Zhang and involve quite different painting skills from the mountains-and-waters type. At the current stage of research, we cannot provide a rigorous evaluation of the result as it demands the expertise of artists.

VII. Conclusions and Future Work

In this paper, we explored how to extract stroke styles of paintings using the mixture of 2-D MHMMs. We have demonstrated that different types of strokes or washes of an artist can be extracted by the mixture modeling approach.

Fig. 12. Comparison between Zhang and Shen's paintings. The above paintings by Zhang are identified as most distinguished from Shen's work.

Several applications can be built upon the trained mixture models: classification of paintings into artists, finding similar or distinguished paintings of one artist to or from another artist's work, and measuring similarity between images of one artist. We have performed experiments on a collection of paintings from Chinese artists. For several reasons, color information and sometimes overall intensity are not helpful for characterizing Chinese paintings. We have shown that based upon high frequency bands of the wavelet transforms of the illuminance component, different styles of strokes can be distinguished. Using the mixture model, experiments on several painting analysis problems have demonstrated promising results.

There is great room for future work in several directions. First, more art-related image analysis tasks should be identified and explored by the developed method. Second, the modeling approach can be advanced. One interesting issue is to automatically determine the complexity of the model. At the current stage, the number of parameters in the mixture model is pre-chosen. Model selection techniques can be applied to adaptively choose the number of parameters according to the style diversity of an artist's paintings. The diversity of an artist's work is itself an interesting aspect to investigate. The developed methodology of the mixture of 2-D MHMMs would provide a basis for forming model selection schemes. Finally, applying this technique to multimodal imaging of art works or images of other domains can be interesting.

Much work needs to be done on the applications of such techniques to assist art historians or users of art image database. Potentially applications can be developed, incorporating image analysis techniques, to allow users to find similar paintings or paintings of similar styles. Applications can also be developed to help art historians to find possible connections among paintings. We will work with the user communities in the future on this.

ACKNOWLEDGMENTS

This work is supported by the US National Science Foundation under Grant No. IIS-0219272, The Pennsylvania State University, the PNC Foundation, and SUN Microsystems under grant EDUD-7824-010456-US. We would also like to acknowledge the comments and constructive suggestions from the reviewers. Conversations with Michael Lesk, Ching-chih Chen, Sally Goldman, and Ruqian Lu have been very helpful. Kurt Grieb offered comments to this manuscript. Ya Zhang provided some assistance in image preparation.

REFERENCES

[1] K. Barnard and D. Forsyth, "Learning the semantics of words and pictures," *Proc. ICCV*, vol. 2, pp. 408-415, 2001.
[2] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 2, pp. 99-113, 1991.
[3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Chapman & Hall, 1984.
[4] G. Celeux and G. Govaert, "Comparison of the mixture and the classification maximum likelihood in cluster analysis," *J. Statist. Comput. Simul.*, vol. 47, pp. 127-146, 1993.
[5] R. Chellappa and A. K. Jain, *Markov Random Fields: Theory and Applications*, Academic Press, 1993.
[6] Y. Chen, *A Machine Learning Approach to Content-based Image Indexing and Retrieval*, Ph.D Thesis, the Pennsylvania State University, 2003.
[7] C.-c. Chen, A.Del Bimbo, G. Amato, N. Boujemaa, P. Bouthemy, J. Kittler, I. Pitas, A. Smeulders, K. Alexander, K. Kiernan, C.-S. Li, H. Wactlar, and J. Z. Wang, "Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials," *DELOS-NSF Reports*, December, 2002.
[8] H. Choi and R. G. Baraniuk, "Image segmentation using wavelet-domain classification," *Proceedings of SPIE*, vol. 3816, pp. 306-320, 1999.
[9] I. Daubechies, *Ten Lectures on Wavelets*, Capital City Press, 1992.
[10] B. Dodson, *Keys to Drawing*, North Light Books, 1990.
[11] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," *Proc. ECCV*, vol. 4, pp. 97-112, 2002.

[12] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2002.
[13] Q. Iqbal and J. K. Aggarwal, "Retrieval by classification of images containing large manmade objects using perceptual grouping," *Pattern Recognition Journal*, vol. 35, no. 7. pp. 1463-1479, 2002.
[14] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075-1088, 2003.
[15] J. Li, R. M. Gray, and R. A. Olshen, "Multiresolution image classification by hierarchical modeling with two dimensional hidden Markov models," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1826-41, August 2000.
[16] J. Li and R. M. Gray, "Context-based multiscale classification of document images using wavelet coefficient distributions," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1604-16, September 2000.
[17] G.J. McLachlan and D. Peel, *Finite Mixture Models*, New York : Wiley, 2000.
[18] M. Pappas, G. Angelopoulos, A. Kadoglou, and I. Pitas, "A database management system for digital archiving of paintings and works of art," *Computers and the History of Art*, vol. 8, no. 2, pp. 15-35, 1999.
[19] S. Ravela and R. Manmatha, "Image retrieval by appearance," *Proc. of SIGIR*, pp. 278-285, Philadelphia, July 1997.
[20] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "WaveCluster: A multi-resolution clustering approach for very large spatial databases" *Proc. of the VLDB Conf.*, pp. 428-439, New York City, August 1998.
[21] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Transactions on Pattern Analysis And Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, 2000.
[22] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.
[23] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, New Jersey, 1995.
[24] J. Z. Wang, J. Li, R. M. Gray, and G. Wiederhold, "Unsupervised multiresolution segmentation for images with low depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 1, pp. 85-90, 2001.
[25] J. Z. Wang, J. Li, S.-c. Lin, "Evaluation strategies for automatic linguistic indexing of pictures," *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, September 2003.
[26] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947-963, 2001.

Jia Li is an assistant professor of Statistics (and Computer Science and Engineering by courtesy) at the Eberly College of Science, The Pennsylvania State University. She received the MSc degree in statistics and a PhD degree in electrical engineering, both from Stanford University (1999). She has been a research staff at Xerox Palo Alto Research Center, Stanford University Computer Science Department, Microsoft-Vxtreme Corp., and Interval Research Corp. Her research interests include statistical learning, data mining, information theory, signal/image processing, and bioinformatics. Her monograph Image Segmentation and Compression Using Hidden Markov Models has been published by Kluwer Academic Publishers and included in the Kluwer Series on Computer Science. She is the author or coauthor of more than 40 journal articles, book chapters, and refereed conference papers.

James Z. Wang received the Summa Cum Laude bachelor's degree in mathematics and computer science from University of Minnesota (1994), the MSc degree in mathematics and the MSc degree in computer science, both from Stanford University (1997), and the PhD degree in medical information sciences from Stanford University Biomedical Informatics Program and Computer Science Database Group (2000). Since 2000, he has been the holder of the endowed PNC Technologies Career Development Professorship and an assistant professor at the School of Information Sciences and Technology and the Department of Computer Science and Engineering at The Pennsylvania State University. He is a member of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials. He has been a visiting scholar at Uppsala University in Sweden, SRI International, IBM Almaden Research Center, and NEC Computer and Communications Research Lab.