Interdisciplinary Research to Advance Digital Imagery Indexing and Retrieval Technologies for Asian Art and Cultural Heritages*

James Z. Wang†
School of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16802
jwang@ist.psu.edu

Jia Li
Department of Statistics
Pennsylvania State University
University Park, PA 16802
jiali@stat.psu.edu

Ching-chih Chen
Graduate School of Library and Information Science
Simmons College
Boston, MA 02155
chen@simmons.edu

ABSTRACT

This paper provides an introduction of our NSF-funded research project on advancing digital imagery technologies for Asian art and cultural heritages. This international collaborative research project aims at developing technologies related to the preservation, retrieval, and dissemination of digital imagery. Researchers in the US, China, and South Korea will collectively investigate and develop technologies for acquiring, browsing, managing, and searching large collections of high quality art images. One of the main research questions the team of US researchers focuses on is the problem of automatic indexing and retrieval of digital art images. Building on the foundation of a successful image retrieval platform, the SIMPLIcity system with the ALIP algorithm, the team is developing techniques to automatically associate linguistic terms with image features for indexing Asian art images. The testbed databases of art images for this research project in the US will begin by using some of the rich image resources of the Emperor and the Chinese Memory Net projects by Ching-chih Chen. This image knowledge base consist of high quality scans, with extensive metadata information including detailed keyword information, as well as comprehensive textual descriptions. The research work aims at demonstrating that (1) modern machine learning and statistical data mining tools are capable of learning from non-structured or semi-structured input data such as human annotations, (2) statistical image modeling techniques can be used in automatic linguistic indexing and concept dictionary building. Finally, we discuss the challenges and the importance for the line of interdisciplinary research work.

Index Terms – Content-based image retrieval, image classification, image annotation, hidden Markov model, computer vision, machine learning, wavelets, Asian art, cultural heritages.

*The URL http://wang.ist.psu.edu provides more information about the project.

†James Z. Wang is also with the Department of Computer Science and Engineering.

1. INTRODUCTION

There is a growing international trend to make information on digital images of art and cultural heritages available to the general public. Most of these images are not currently accessible. The Internet and the Web are excellent medium for distributing imagery and other important work [7]. Some amateurs have already started to scan art images from book publication free of copyright. Yet, the collection of scanned art images found on the Web is of small quantity and of amateur quality. Because the scanned images are located in so many places on the Web, it is often impossible to locate the images when they are needed. It is important to develop technologies that provide content-based search of distributed databases of art and cultural heritages.

The main goal of our National Science Foundation (NSF) supported research is to advance information technologies related to the preservation, retrieval, and dissemination of digital imagery for Asian art and cultural heritages. The project, based on the development and use of the Stanford Chicana Art image database since 1995 and the collaborative work [2] with Chinese Memory Net (CMNet) [3], this project will start by further development and testing work on the rich and substantial image and video knowledge bases on the world renowned terracotta warriors and horses of the First Emperor of China [4], developed by Ching-chih Chen of Simmons College. The extensive metadata with detailed key words as well as descriptive annotations will be invaluable to this project since they are very labor-intensive to create and require subject knowledge. As additional high quality ancient art images, such as those related to ancient Asian paintings and calligraphies, become available, this project will extend to include those contents.

1.1 Related work on indexing images

Many content-based image retrieval (CBIR) systems have been developed since the early 1990s. A recent article published by Smeulders et al. reviewed more than 200 references in this ever changing field [9]. Readers are referred to that article and some additional references [14, 12, 5] for more information.

Most of the CBIR projects aimed at general-purpose image indexing and retrieval systems focusing on searching images visually similar to the query image or a query sketch. They do not have the capability to assign comprehensive textual description automatically to pictures, i.e., linguistic indexing, because of the great difficulties in recognizing a large number of objects. However, this function is essential for linking images to text and consequently broadening the possible usages of an image database.

1.2 Our statistical approach

A growing trend in the field of image retrieval is to linguistically index images using computer programs relying on statistical classification methods. The Stanford SIMPLIcity (Semantics-sensitive Integrated Matching for Picture LIbraries) system [11] uses manually-defined statistical classification methods to classify the images into rough semantic classes, such as textured-nontextured, graph-photograph. Potentially, the categorization enhances retrieval by permitting semantically-adaptive searching methods and narrowing down the searching range in a database. The approach is limited because these classification methods are problem specific and must be manually developed and coded. We have applied the SIMPLIcity system to various areas including the Emperor image database [2].

A recent work on associating images explicitly with words is that of University of California at Berkeley [1], in which a hierarchical clustering model incorporating image features and text information is established to organize images in a database.

In the recent work carried out at Penn State University [10], categories of images, each corresponding to a concept, are profiled by statistical models, in particular, the 2-dimensional multi-resolution hidden Markov model (2-D MHMM) [8]. The pictorial information of each image is summarized by a collection of feature vectors extracted at multiple resolutions and spatially arranged on a pyramid grid. The 2-D MHMM fitted to each image category plays the role of extracting representative information about the category. In particular, a 2-D MHMM summarizes two types of information: clusters of feature vectors at multiple resolutions and the spatial relation between the clusters, both across and within resolutions. As the estimation of a 2-D MHMM is done separately for each category, a new category of images added to the database can be profiled without repeating computation involved with learning from the existed categories. Since each image category in the training set is manually annotated, a mapping between profiling 2-D MHMMs and sets of words can be established. For a test image, feature vectors on the pyramid grid are computed. Consider the collection of the feature vectors as an instance of a spatial statistical model. The likelihood of this instance being generated by each profiling 2-D MHMM is computed. To annotate the image, words are selected from those in the text description of the categories yielding highest likelihoods. The research resulted in the ALIP (Automatic Linguistic Indexing of Pictures) system which is capable of building a dictionary of 600 concepts automatically.

In the coming years, we plan to further develop the ALIP system for the purpose of indexing art images. The manually-annotated Emperor multimedia knowledge base developed by Chen as a part of CMNet with labor-intensive manual annotations will serve as an important testbed for this research. Computer programs will attempt to learn to build a knowledge base from the existing metadata. Potentially, such automatically-generated knowledge bases can be used by computers to annotate images of similar semantic content.

2. SIMPLICITY AND ART

Figure 1: Query results of the Chicana Art image search engine developed by Wang for the Stanford University Art Library. The first image is the query image. Some paintings with similar styles are retrieved.

Our development of image indexing and retrieval systems [11] started as early as 1995 at Stanford University. The first project, conducted by J. Z. Wang et al., was initiated by the Stanford University Libraries and later funded by IBM QBIC, NEC C&C Research Labs, SRI International, and a research grant from US NSF. The goal was to design and implement a computer system capable of indexing and retrieving large collections of digitized multimedia data available in the libraries based on the media contents. At the time, it seemed reasonable that one should discover the solution to the image retrieval problem during the project. Experience has certainly demonstrated how far we are as yet from solving this basic problem. The problem is challenging because of the large size of the database, the difficulty of understanding images, both by people and computers, the difficulty of formulating a query, and the problem of evaluating the results.

After the first system, the WBIIS (Wavelet-Based Image Indexing System) [14], and its application to the Chicana Art database (Figure 1), Wang and Li realized the importance of region-based indexing in retrieving arts. A region-based retrieval system applies image segmentation [13] to decompose an image into regions, which correspond to objects if the decomposition is ideal. The object-level representation is intended to be close to the perception of the human visual system. However, image segmentation is nearly as difficult as image understanding because the images are 2-D projections of 3-D objects and computers are not trained in the 3-D world the way human beings are.

In 1999, Wang and Li developed the SIMPLIcity system [12]. As in other region-based retrieval systems, an image is represented by a set of regions, roughly corresponding to objects, which are characterized by color, texture, shape, and location. The system classifies images into semantic categories, such as textured-nontextured, graph-photograph. Potentially, the categorization enhances retrieval by permit-

ting semantically-adaptive searching methods and narrowing down the searching range in a database. A measure for the overall similarity between images is developed using a region-matching scheme that integrates properties of all the regions in the images. Compared with retrieval based on individual regions, the overall similarity approach (1) reduces the adverse effect of inaccurate segmentation, (2) helps to clarify the semantics of a particular region, and (3) enables a *simple* querying interface for region-based image retrieval systems. The application of SIMPLIcity to several databases, including a database of about 200,000 general-purpose images, has demonstrated that our system performs significantly better and faster than existing ones. The system is fairly robust to image alterations.

The SIMPLIcity system has been applied to many areas, demonstrated by the fact that more than 40 institutions including universities, government agencies, and NASA JPL, have obtained the research license of the software. Recently, Chen and Wang successfully applied the SIMPLIcity system to the problem of searching the art images of The First Emperor of China's terracotta warriors, horses and other related art objects as a collaborative effort of CMNet [2]. The preliminary results are most relevant to this project and thus deserve some brief discussion in the following section.

2.1 Application of SIMPLIcity in art history and archaeology

In addition to the popular interactive videodisc and later the multimedia CD-ROM products both published by the Voyager Co., Chen's Emperor Project supported by the National Endowment for the Humanities has also a very valuable crude database which provides significant metadata information on each of the 5,000 most significant images [3]. For the NSF/IDLP's CMNet, Chen has further modified and expanded this database to a dynamic image knowledge base with comprehensive metadata information as shown in Figure 2.

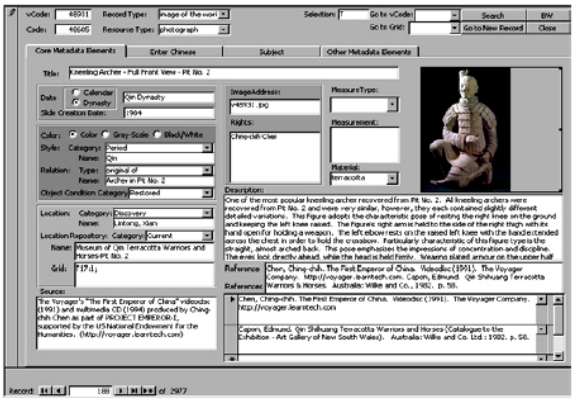

Figure 2: A part of a typical record in the comprehensive Emperor database.

This database is developed under the Microsoft Access environment and can also be easily ported to other relational databases such as Oracle. The database engine provides fast search results on individual indexing attributes.

Although all Emperor images were scanned in a very high resolution of 1200 dpi, they are saved in 5 different derivatives for other research and development purposes. Our initial collaboration uses only the smallest icon images with the SIMPLIcity technology, and the results were already quite good. That shows the high robustness of the mathematical and statistical algorithms underlying the system.

Figure 3: SIMPLIcity shows a random selection of images from the Emperor database.

The SIMPLIcity system allows the user to interact and search the Emperor database in different ways. Figure 3 shows the "Random" mode of the system which gives user a random selection of images from the database. The user may choose one of the images from the selection as a query image to find similar images from the database. Figures 4 and 5 show the search results on some sample images.

Figure 4: SIMPLIcity search result. The upper-left corner image is the query image the user selected.

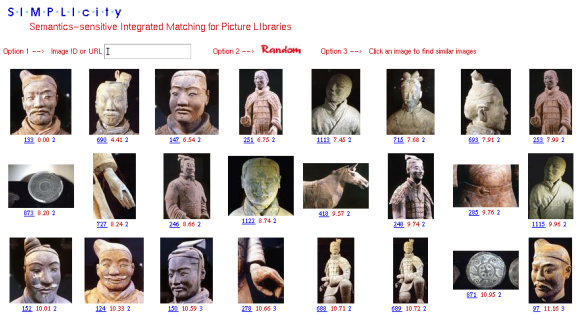

Figure 5: Another SIMPLIcity search result. The upper-left corner image is the query image the user selected.

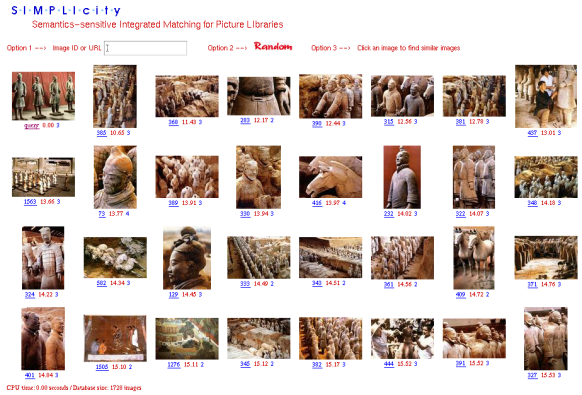

Figure 6: SIMPLIcity is capable of quickly locating the similar images to a query image from the Web. The query image, shown as the upper-left corner image, was downloaded from http://www.unc.edu/courses/hist033 .

Because of the fast image segmentation and region-based feature indexing speed of the SIMPLIcity system, it permits the user to search on a query image from anywhere on the Internet in real time. The user may enter the URL of the query image in the search field. The server downloads the image from that URL, extracts the features from the image, and compares the query image to all images in the database. Typically, it takes a couple of seconds to perform these operations using a Pentium PC based server.

The user is allowed to draw the query image using a JAVA interface. However, because of the difficulty in drawing textures, the time for a user to formulate a query using the drawing applet can be long. We are investigating better drawing interfaces.

3. ALIP

Recently, the team at Penn State University developed the ALIP system [10]. The main motivation for this work is to empower the computer with semantic learning capabilities so that linguistic indexing of pictures may be possible. A picture is worth a thousand words. As human beings, we are able to tell a story from a picture based on what we have seen and what we have been taught. A 3-year old child is capable of building models of a substantial number of concepts and recognizing them using the learned models stored in her brain. Can a computer program learn a large collection of semantic concepts from 2-D or 3-D images, build models about these concepts, and recognize them based on these models? This is the question the ALIP work attempts to address.

Automatic linguistic indexing of pictures is essentially important to content-based image retrieval and computer object recognition. It can potentially be applied to many areas including biomedicine, commerce, the military, education, digital libraries, and Web searching. Decades of research has shown that designing a generic computer algorithm that can learn concepts from images and automatically translate the content of images to linguistic terms is highly difficult. Much success has been achieved in recognizing a relatively small set of objects or concepts within specific domains. There is a rich resource of prior work in the fields of computer vision, pattern recognition, and their applications [6].

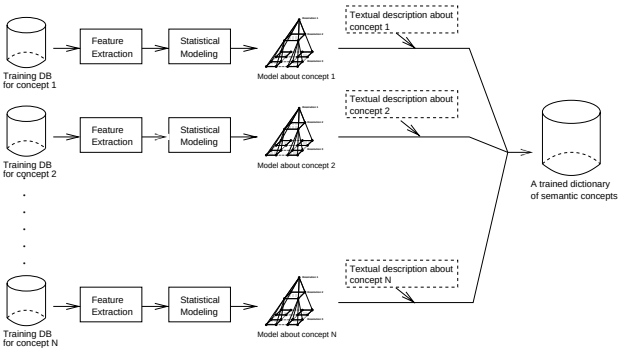

Figure 7: The architecture of the statistical modeling process.

Readers are referred to [10] for technical details of the ALIP system and the evaluation of its performance. The system has three components: the learning component, the statistical comparison component, and the linguistic indexing component. Figures 7 and 8 illustrate the flow of the statistical modeling and the image indexing processes of the system.

In the learning phase, we train the computer system using categories of images. For each concept category, we provide the system with a set of images depicting the concept. The computer builds a multi-scale statistical model after analyzing the features extracted from the images. To validate the methods, the ALIP was implemented and tested with a general-purpose image database including about $60,000$ photographs. These images are stored in JPEG format with size 384×256 or 256×384. The system is written in the C programming language and compiled on two UNIX platforms: LINUX and Solaris.

We conducted experiments on learning-based linguistic indexing with a large number of concepts. The ALIP system was trained using a subset of $60,000$ photographs which are based on 600 CD-ROMs published by COREL Corp. Typically, each COREL CD-ROM of about 100 images represent one distinct topic of interest. For our experiment, the dictionary of concepts contains all 600 concepts, each associated with one CD-ROM of images.

We manually assigned a set of keywords to describe each CD-ROM collection of 100 photographs. The semantic descriptions of these collections of images range from as simple or low-level as "mushrooms" and "flowers" to as complex or high-level as "England, landscape, mountain, lake, European, people, historical building" and "battle, rural, people, guard, fight, grass". On average, 3.6 keywords are used to describe the content of each of the 600 concept categories. It took the authors approximately 10 hours to annotate these categories.

After running the learning component of the ALIP system, a dictionary or knowledge base of 600 concepts is built by the computer automatically. In the image indexing phase, the computer compares the features in an un-annotated image

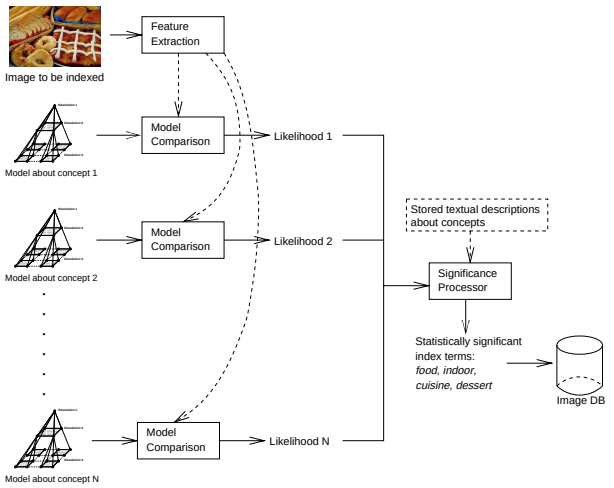

Figure 8: The architecture of the statistical linguistic indexing process.

with the stored concept models. Statistical likelihood is used to indicate whether an image resembles a given concept. Finally, the statistical significance of each possible indexing keyword is assessed to determine the annotation keywords for each image.

3.1 Potential applications of ALIP in art

For art image databases, keyword-based manual annotations are simply too labor-intensive and requires substantial subject knowledge which generally are not possible for the technologists to create. It is also often too expensive to annotate a very large scale art image database by hand. This is why Chen's Emperor image knowledge base is so significant for this research. Essentially we have a ready made scholarly knowledge base with all the required information needed to apply and test our ALIP system. This technology-content mix is a real necessity to enable us to move our research from simple commercial images to real art history and archaeology images with great research and educational potential.

One of the main benefits of using the Emperor image database is its comprehensive structured and semi-structured metadata. This allows the computer system to learn concepts ranging from very simple ones to very complex ones. The metadata of the database is carefully prepared. We expect the system to be able to capture some domain-specific concepts based on the expert annotation. We aim at demonstrating that modern machine learning and statistical data mining tools are capable of building domain concept dictionary automatically and use the learned models in automatic linguistic indexing of images.

We also hope to apply the ALIP system to Asian painting image database when available. We are in the process of obtaining some digital images of the most significant ancient Chinese paintings for this research. When they are available, the ALIP system will be used to analyze these paintings. Potentially, the computer system will capture some of the most important concepts in the domain of ancient Asian paintings. We expect that the computer system will be able to answer questions like: "What are the most important clues for art historians to tell if a painting is from the Tang Dynasty or the Qing Dynasty?" and "What features are the discriminative features for Chinese landscape paintings?"

These are very challenging problems, which will be further elaborated in the next section.

4. DISCUSSIONS

Research on digital imagery technologies for art and cultural heritages is critically important for its great potential in further advancing related sciences and engineering, its relevance of arts and cultural heritages to education at various levels, and its role in promoting cultural understanding.

Research on art images is likely to shed light into many other image-related research fields such as computer vision and efficient transmission of images. When an artist paints a picture, he/she is not simply copying what is seen. Putting aside the aspects of expressing imagination and mood, even when the intention is to display the real world accurately, special techniques are exploited to provide viewers a sense of three dimensions and subtle lighting using a flat canvas and painting media with limited colors and shades. It took many genius artists centuries to develop these techniques. One such example is the use of composite colors to achieve high contrast, a technique pioneering impressionism painters Monet and Renoir discussed, studied, and mastered to brighten oil painting. By using a great variety of textures, Van Gogh even conveys through his paintings a touching feeling of objects. Art pictures are thus records for how the real world is captured in images by artists, including beyond any doubt the most talented people who understand the link between them. Therefore, art pictures are of great values in their own right for research on images. Applying modern computing techniques to analyze them will gain insights for general-purpose image archiving, distributing, and intelligent automatic information extracting.

Art is a crucial part of education for children and the general public. As a treasure of the human culture, it provides people inspiration, imagination, and proud. Art work records history. Paintings have existed since the dawn of the civilization, as a sharp contrast to any other imagery technology, which arose in the recent few hundred years. Paintings in a particular era show the social structure, the way people normally lived, the fashion and entertainment, and sometimes the technological level of the time. Paintings reflect artists' imagination, mood, personality, belief, and even social attitude. They are a form of high-level creation, not a sheer craft. This is why an important branch of Chinese painting is referred to as 'write about the essence'. As a fruit of the highly elaborated human intelligence, it is not surprising art has inspired numerous people, enhanced their lives, and made them confident in themselves and respectful to others. Not every child has the opportunity to visit museums or read many art books. Research on digital imagery techniques for art pictures will make art work much easier to access and to study.

Exposing the general public to art from different cultures will increase understanding and appreciation between people with different cultural backgrounds. Art work can be appreciated relatively easily across cultures since no spe-

cialized capabilities, e.g., language, are needed. The intrinsic beauty in art also leads people to a relatively open mind towards different traditions. Nowadays, with the ever increasing communication between people all over the world, it is crucial that people respect cultural diversities and learn from each other. Prejudices often come from misunderstanding, or unwillingness to understand. Art is an excellent cultural representative in the sense of helping people to look at other cultures objectively. Modern digital technologies have made it a reality to exhibit large collections of art work from multiple cultures. Since an enormous amount of art work has been created, both storage and distribution raise many challenges. Further advancing digital technologies for archiving and distributing art work is of great importance.

5. CONCLUSIONS

In this paper, we gave an introduction of our NSF-funded research project on advancing digital imagery technologies for Asian art and cultural heritages. We have provided the overall designs of the the SIMPLIcity content-based image retrieval system and the ALIP automatic linguistic indexing of pictures system. We have described the collaborative effort in using the comprehensive Emperor's image knowledge base of Chen's CMNet project. We have illustrated how the SIMPLIcity system is used in searching art images. Potentially, the application of the ALIP system in using the Emperor's rich annotations and keywords will demonstrate that statistical learning and data mining methods can be used by computers to automatically learn domain-specific knowledge for the purpose of intelligent image annotation. This application will be applied to other art topics.

6. ACKNOWLEDGMENTS

The SIMPLIcity work was supported in part by the US National Science Foundation (NSF) under Grant No. IIS-9817511 and Stanford University. The research and development work related to EMPEROR was supported by the National Endowment for the Humanities for PROJECT EMPEROR-I and NSF/IDLP for Chinese Memory Net under Grant No. IIS-9905883. This work is supported primarily by The Pennsylvania State University, the NSF under Grant No. IIS-0219272, the PNC Foundation, and SUN Microsystems under grant EDUD-7824-010456-US. Our other collaborators will include Zuoquan Lin, Ruqian Lu, Kyu-Young Whang, and Gio Wiederhold. Conversations with Michael Lesk have been very helpful. The copyright of the original Chicana Art images shown in this paper belongs to Stanford University. The copyright of the original Emperor images shown in this paper belongs to Ching-chih Chen of Simmons College.

7. REFERENCES

[1] K. Barnard, D. Forsyth, "Learning the Semantics of Words and Pictures," In *Proc. ICCV*, 2:408-415, 2001.

[2] C.-C. Chen, J. Z. Wang, "Large-scale Emperor digital library and semantics-sensitive region-based retrieval," In *Proc. International Conference on Digital Library – IT Opportunities and Challenges in the New Millennium*, 454-462, Beijing:National Library of China, July 9-11, 2002.

[3] C.-C. Chen, "Chinese Memory Net (CMNet): A model for collaborative global digital library development," In *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*, C.-C. Chen (ed.), 21-32, Beijing:Tsinghua University Press, 2001.

[4] C.-C. Chen, "Multimedia and the First Emperor of China: Moving knowledge base," *Multimedia Today (IBM)*, 2(2):68-71, April 1994.

[5] Y. Chen, J. Z. Wang, "A region-based fuzzy feature matching approach to content-based image retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(9):1252-1267, 2002.

[6] D. A. Forsyth, J. Ponce, *Computer Vision: A Modern Approach,* Prentice Hall, 2002.

[7] M. Lesk, *Practical Digital Libraries: Books, Bytes, and Bucks*, Morgan Kaufmann Publishers, 1997.

[8] J. Li, R. M. Gray, R. A. Olshen, "Multiresolution image classification by hierarchical modeling with two dimensional hidden Markov models," *IEEE Trans. on Information Theory,* 46(5):1826-1841, August 2000.

[9] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. on Pattern Analysis And Machine Intelligence,* 22(12):1349-1380, 2000.

[10] J. Z. Wang, J. Li, "Learning-based linguistic indexing of pictures with 2-D MHMMs," In *Proc. ACM Multimedia*, Juan Les Pins, France, ACM, December 2002.

[11] J. Z. Wang, *Integrated Region-based Image Retrieval,* Kluwer Academic Publishers, Dordrecht, 2001.

[12] J. Z. Wang, J. Li, G. Wiederhold, "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 23(9):947-963, 2001.

[13] J. Z. Wang, J. Li, R. M. Gray, G. Wiederhold, "Unsupervised multiresolution segmentation for images with low depth of field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(1):85-91, 2000.

[14] J. Z. Wang, G. Wiederhold, O. Firschein, X. W. Sha, "Content-based image indexing and searching using Daubechies' wavelets," *Int. J. of Digital Libraries(IJODL)*, 1(4):311-328, Springer-Verlag, 1998.