BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

TEXTURE- AND STRUCTURE-BASED IMAGE REPRESENTATION WITH APPLICATIONS TO IMAGE RETRIEVAL AND COMPRESSION

by

ZHIHUA HE

B.S., University of Electronic Science and Technology of China, 1999
M.S., Drexel University, 2002

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2007

UMI Number: 3246606

INFORMATION TO USERS

Approved by

First Reader ______________________________
Maja Bystrom, Ph.D.
Associate Professor of Electrical and Computer Engineering

Second Reader ______________________________
W. Clem Karl, Ph.D.
Professor of Electrical and Computer Engineering and Professor of Biomedical Engineering

Third Reader ______________________________
Janusz Konrad, Ph.D.
Associate Professor of Electrical and Computer Engineering

Fourth Reader ______________________________
Stan Sclaroff, Ph.D.
Associate Professor of Computer Science

Acknowledgments

I am usually not good at writing acknowledgment, one reason is that I am never good at writing, and another reason is that I believe action speaks more for words. However, when I look back over all these years I spent in schools, I do realize that I owe lots of thanks to people along the way.

First, I would like to thank my parents, ShuiBao He and Yong Xu. They have provided me with a stable home, without which I can never dream of getting advanced education; especially thanks for my father, who gave me a dream of being a scientist from my early childhood. I definitely want to thank my brother, Zhipeng He, who takes the responsibility to take care of the family while I am away from home.

Then, a special thank goes to my PhD adviser, Dr. Maja Bystrom. Without her continuous support, I will not be able to come here for my PhD research. Without her guidance and patience, I can hardly imagine how I could get through the PhD research alone. I will not forget the warmth brought by her in those cold winters. I also want to thank Dr. Arkady Kopansky, my first office-mate here, who has been always willing to help and answer the questions.

Here, I want to thank my committee members: Dr. Stan Sclaroff, Dr. Janusz Konrad and Dr. Clem Karl. Thank you for your careful reviews and valuable comments.

In the mean time I would also like to express my gratitude to all my teachers from primary school to college. You have led me step by step into the science world, and have shown me how to approach problems from different aspects.

Finally, I want to express my gratitude to all my friends, life will be so lonely and so boring without you guys! It is really hard for me to name all of you here, but I want to let you all know that I really appreciate your love and care, and I will always be willing to share mine with you. Below I want share my special thanks to Fang, Doreen, Violet, Shuang and Ling. Thank you for being there with me always.

TEXTURE- AND STRUCTURE-BASED IMAGE REPRESENTATION WITH APPLICATIONS TO IMAGE RETRIEVAL AND COMPRESSION

(Order No.)

ZHIHUA HE

Boston University, College of Engineering, 2007

Major Professor: Maja Bystrom, Ph.D.,
Department of Electrical and Computer Engineering

ABSTRACT

The design of an efficient image representation methods using small numbers of features can facilitate image processing tasks such as compression of images and content-based retrieval of images from databases. In this dissertation, three methods for capturing and concisely representing two distinguishing characteristics of images, namely texture and structure, are developed. Applications of these compact representations of image characteristics to image compression as well as retrieval of images and hand-sketches of images from databases are given and performance is compared with other compression and retrieval methods.

The first method to be introduced is a directional, hidden-Markov-model-based method for succinctly describing image texture using a small number of features. This method employs the well known, multi-scale contourlet and steerable-pyramid transforms to isolate in different subbands the edges that comprise the image texture. Statistical inter- and intra-subband dependencies are captured via hidden Markov models, and model parameters are used to represent texture in small feature sets. Application of this method to content-based retrieval of images with homogeneous textures from database is shown. At the similar computation cost, about 10% higher retrieval rates over comparable methods are

demonstrated; when approximately one third fewer features are used, similar retrieval rates can be obtained using the proposed method.

A method for concisely describing large image structures, that is, significant image edges, is then proposed. This method decomposes an image using the contourlet transform into directional subbands which contain edges of different orientations. Each subband is then projected onto its associated primary and orthogonal directions and the resulting projections are filtered and then modeled using piece-wise linear approximations or Gaussian mixture models. The model parameters then form the concise feature sets used to represent the image's structure. An application of this image-representation method to retrieval of images from databases based on users' sketches of the images is shown. An retrieval rate increase of 13% using the proposed method is demonstrated over a current spatial-histogram-based method.

Finally, a new multi-scale curve representation framework, the chordlet, is constructed for succinct curve-based image structure representation. This framework can be viewed as an extension to curves of the well known beamlet transform, a multi-scale line representation system. In this dissertation, the representation efficiency, in terms of bits versus distortion, of the chordlet transform is compared with that of the beamlet transform. An algorithm for performing a fast chordlet transform has been developed. A chordlet-based coding system is constructed for application of the chordlet transform to compression of image shapes. By using the proposed method increased compression is obtained at lower distortion when compared with two well known methods.

Contents

List of Figures

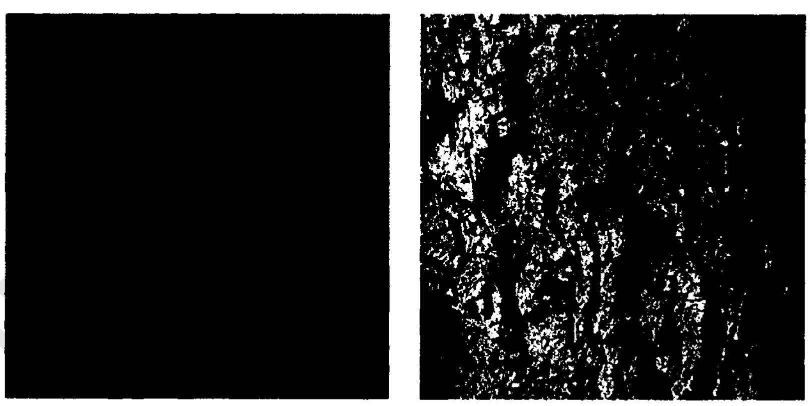
Chapter 1

Introduction

Image representation serves as an integral part of image processing algorithms. For different applications, designing or selecting an appropriate representation for image data plays an important role. While image representation is a broad research field, in this dissertation, we are interested in efficient representation of primary image features. More specifically, texture representation and structure representation are studied. All of the work included in this dissertation is based on multi-scale frameworks. To help the readers have a better understanding of our work, Chapter 2 gives an overview of multi-scale representations. In particular, the directional multi-scale representations are discussed in more details, which are used throughout this dissertation.

Examples of texture images are given in Figs. 1·1 (a)-(b). In texture representation, the key is to find an accurate and efficient description of the texture information in an image. This representation could be used for texture-related applications such as texture classification and texture retrieval. For example, for an image with checker-board patterns, one texture representation could be simply text, e.g., *'checker-board texture'*. Another texture representation could be the histogram of the image, which characterizes the statistical distribution of the colors/intensities. Furthermore, different models can be adopted for the characterization of color/intensity distributions. For this example, if the user would like to find the image with the similar textures as this checker-board image in an image database, which is known as texture-based image retrieval, then the text-based representation might be applicable for retrieval in small database. However, when large image database is considered, text annotation of every image in the database becomes very expensive, thus, the histogram-based representation might be more appropriate, since the representation can be automatically generated for all the images. On the other hand, the histogram-based representation requires the storage of the entire histogram, and the histogram only characterizes the frequencies of colors/intensities in its corresponding image and discards geometry information, therefore, it might be a poor descriptor for images with similar textures. Thus, more sophisticate statistical modeling of the color/intensity distribution may offer a more flexible and robust representation for texture-based retrieval applications.

A strict definition of texture is debatable because of its observer-dependent nature for different applications (Wang et al., 1998). In general, two classes of methods, namely, the structural-based and statistical-based methods, are used for texture representation (Long et al., 2003). Structural-based representation is typically used for textures with strong and regular patterns. On the other hand, statistical-based methods are used more widely. Assuming textures are generated from certain statistical sources, statistical models, for which rich mathematical frameworks are available, can be constructed to capture texture information in a small number of parameters. However, depending on the construction of statistical models, structural information of textures might be lost or reduced to certain extent.



(a) (b)

Figure 1·1: Example of texture images. (a) Image *Tile.* (b) Image *Bark.*

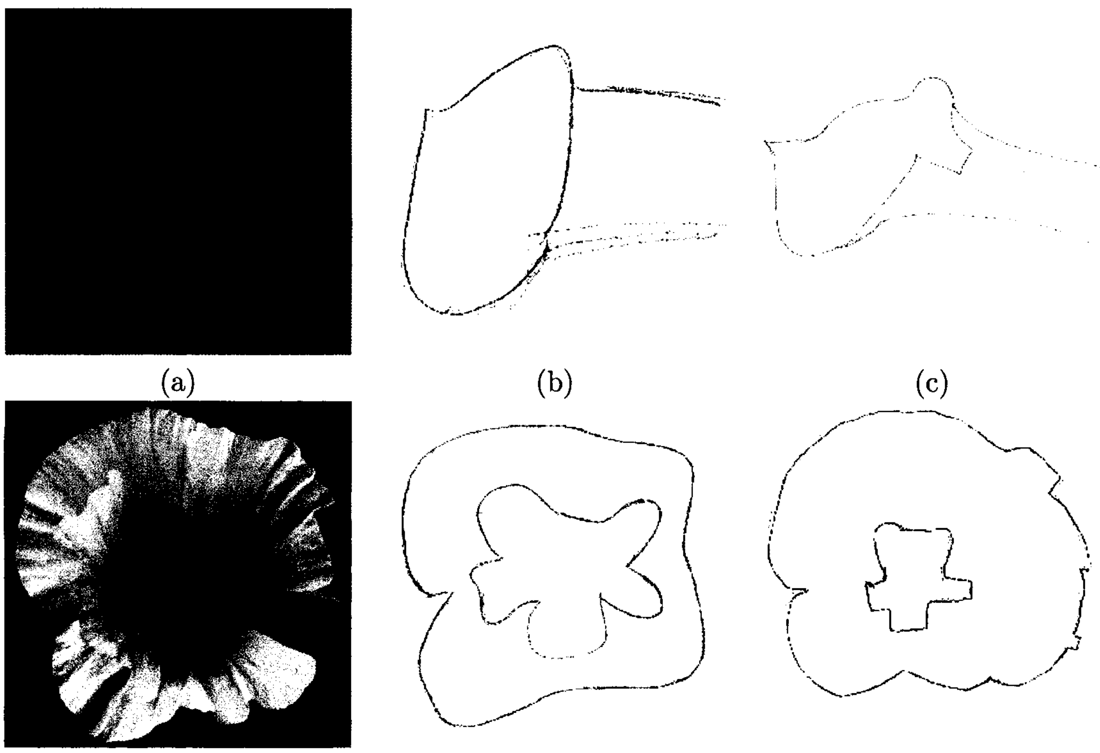
Many texture feature models are derived from an early work of Tamura, et al. (Tamura et al., 1978), which models texture features based on human texture perceptions using coarseness, contrast, directionality, linelikeness, regularity, and roughness. In recent years, with the wide applications of multi-scale analysis in image processing, active studies (Randen and Husoy, 1999; Manjunath et al., 2000; Wouwer et al., 1999; Crouse et al., 1998) have been carried out on texture models in the multi-scale spatial-frequency domains, in particular, the wavelet domain. The popularity of multi-scale wavelet decomposition in image analysis is partly supported by the physiological studies of the human visual system (HVS) (Valois and Valois, 1988), and facilitated by the success of wavelets in still image compression in JPEG-2000 (Joint Photographic Experts Group Homepage, 2005). In (Randen and Husoy, 1999), Randen and John gave a comparative study of texture feature extractions using different filter banks, with local energy from each filter bank used as features. The role of different filter banks is compared by using the same similarity measure for corresponding texture features in retrieval in (Randen and Husoy, 1999). The first- and second-order statistics of the wavelet coefficients are employed by Wouwer, et al. in (Wouwer et al., 1999) for texture modeling. A texture descriptor based on the Gabor decomposition is described in (Manjunath et al., 2000), where the structuredness of the textures are captured by a perceptual browsing component and the statistics of the texture is captured by a similarity retrieval component.

While texture models based on local energy and basic statistics in the wavelet domain yield reasonable retrieval performance in texture-based retrieval applications, more sophisticated statistical models in the wavelet domain are developed to capture both statistical and structural properties of textures. In particular, wavelet-based hidden Markov models (HMMs) are widely used in texture modeling (Crouse et al., 1998; Do and Vetterli, 2002a; Fan and Xia, 2003; Po, 2003). In (Crouse et al., 1998), Crouse, et al. exploited cross-scale hidden Markov dependencies in the estimation of statistical model parameters for feature extraction in the wavelet domain, and suggested possible further extensions utilizing cross-band and in-band (spatial) dependencies. In (Do and Vetterli, 2002a), Do and Vetterli adopted a vector HMM model for characterization of textures in the steerable pyramid wavelet domain, investigating both cross-scale and cross-band dependencies. Fan and Xia (Fan and Xia, 2003) have proposed a HMM model for texture retrieval and synthesis also considering cross-band correlation. Recently, Po and Do (Po, 2003) have studied statistical modeling in the contourlet domain (Do and Vetterli, 2003), and exploited cross-scale dependencies for their HMM approach to the texture retrieval problem. In Chapter 3, we investigate the importance of different dependencies in constructions of contourlet-based HMMs, along with possible combination of color information for texture representation. The application to the texture-based image retrieval is studied.

The high-level visual contents of an image, including objects' shapes and their spatial relationships, can be characterized in shape and spatial features. Shape and spatial feature extraction is often object-based, i.e., an image is usually first classified/segmented into regions/objects before feature extraction. Two classes of methods have been used for shape extraction, namely, boundary-based approaches and region-based approaches (Long et al., 2003). Boundary-based approaches extract features based on object contours; this includes approaches such as Fourier-based shape descriptors (Persoon and Fu, 1977; Kauppinen et al., 1995), and curvature-based shape descriptors (Abbasi et al., 1999). The classical region-based approaches focus on the statistical moments of the object region (Hu, 1962; Yang and Albregtsen, 1994). Spatial features representing the spatial relationships between objects can be modeled using 2-D strings (Chang et al., 1987; Lee et al., 1992) or symbolic images (Gudivada and Raghavan, 1995). Most of the methods for shape and spatial feature extraction rely on image segmentation; however image segmentation is another active research field, and accurate segmentation is in general difficult to achieve. Therefore, applications relying on robust shape features and spatial feature extraction are limited as well. On the other hand, structure representation, which includes shape and spatial information, can be adopted with more flexibility and robustness in different image applications.

Many of approaches to structure representation focus on closed shape description (Latecki et al., 2000; Grigorescu and Petkov, 2003; Abbasi et al., 1999; Persoon and Fu, 1977).

While these results are important for many applications, the work presented in this dissertation focuses on representation of the edges/boundaries in an image without restriction to closed shape. The edges/boundaries in an image are usually approximated by line segments and curve segments. Accordingly, two structure representation frameworks are constructed in this work, namely, directional-projection-based structure representation (DP-SR) and curve representation via a chordlet framework. The DP-SR is constructed based on the multi-scale contourlet pyramid, it captures line-segment information through projection of the edges along different orientations, and summarizes the line-approximated edges in a set of projection profiles through profile modelings. Projection-based methods such as the Hough transform, can efficiently capture line features (Tipwai and Madarasmi, 2002; Franti et al., 2000; Fung et al., 1996). However, the Hough transform treats points on integration lines equally, and thus reduces the spatial information of the shapes. Alternatively, Kadyrov and Petrou (Kadyrov and Petrou, 2001) proposed a trace transform as the generalization of the Hough transform, where selected functionals along the straight lines are utilized and some invariance to translation, rotation and scaling is achieved. In (Abdel-Mottaleb, 2000), histograms of edge segments in four orientations of a block of pixels are used as the shape representation. Tang and Wang (Tang and Wang, 2004) approached the face sketch recognition problem by first transforming face photos into pseudo-sketches, and then comparing the query sketches with pseudo-sketches in a database using an eigenface representation method. In (Sclaroff, 1997), Sclaroff has described a method deformable shape-based image retrieval, where the modal matching technical is used. In DP-SR, through a contourlet transform, edges with different orientations are captured in directional subbands at various scales. Following the projection of each subband into a pair of orthogonal profiles, the shape and spatial information is compactly captured by identifying the primary structures of each 1-D projection profile. In Chapter 4, we first introduce the proposed directional projection based structure representation DP-SR, based on which a piece-wise peak-valley approximation-based model DP-SR-I is developed. Then we describe a non-linear Gaussian mixture approximation-based model DP-SR-II. The advantages and disadvantages of the proposed models are discussed, and a potential application of DP-SR to image retrieval based on sketches is described with simulations performed. Examples of sketches and their original images are given in Figs. 1·2 (a)-(f).



(a) (b) (c)

(d) (e) (f)

Figure 1·2: An example of sketches and corresponding images. (a) *Flower 1.* (b)-(c) Two sketches of *Flower 1.* (d) *Flower 2.* (e)-(f) Two sketches of *Flower 2.*

Among the multi-scale representations, it has been shown in (Donoho et al., 1998) that wavelets are optimal in capturing point singularities, thus, they provide good representations for data with point discontinuities. To provide a good representations for line segments, which are crucial in 2D image representations, the curvelet (Candes and Donoho, 1999a), contourlet (Do and Vetterli, 2003) and beamlet (Donoho and Huo, 2002) have been constructed. Chapter 2 will give introduction to these transforms, and a brief review of them is given below. In the curvelet (Candes and Donoho, 1999a), line singularities, are captured through a double transform, i.e., the wavelet transform and ridgelet transform, where the wavelet transform captures the point singularities, whose accumulation, line singularities are captured through a ridgelet transform. The construction of contourlet (Do and Vetterli, 2003) can be viewed as a discrete approximation of the curvelet. A double filter bank, Laplacian pyramid and directional filter banks, framework is adopted. Besides the curvelet and contourlet, the beamlet (Donoho and Huo, 2002) offers a straight forward and flexible framework for line segment representation. In the beamlet framework, a dictionary of line segments at different scales and different locations is defined; the elements in the dictionary are used to represent arbitrary line segments in the image. In a summary, the curvelet, contourlet and beamlet all focus on the line-segment representation as approximation to curves.

In this dissertation, structure representation using curve segments is studied through the construction of the chordlet framework, a hierarchical multi-scale curve representation. The chordlet can be viewed as the curve extension of beamlet. Through a representation efficiency model constructed in Section 5.3, the chordlet-based representation is first shown to be theoretically efficient as compared to beamlet-based representation. A fast chordlet transform algorithm, using the combination of direct and FFT-based chordlet transform, is proposed in Section 5.4 and its performance has been evaluated. For the application to thresholding-based representation, a chordlet coefficient significance model is proposed in Section 5.5 to help the user determine the significance level for the chordlet coefficients in representation system. A simple thresholding-based chordlet representation system is then described based on this significance model. To resolve the in-band and cross-scale correlation between the chordlet elements in chordlet-based representation, two chordlet coding systems are constructed in Section 5.7 using a rate-distortion model. The application of the proposed chordlet coding systems to shape structure representation/compression is studied and compared with beamlet-based coding system. The derivation and discussion of coding efficiency, fast implementation, and coefficient significance in Chapter 5 suggest that the proposed chordlet provides a good curve representation for image analysis. The simulation