CHAPTER 1

INTRODUCTION

1.1 Motivations

“Vision is the process of discovering from images what is present in the world and where it is” (Marr [88], p. 3). Due to the apparent simplicity of the action of seeing, however, the underlying difficulties of visual information processing had not been realized until Marr’s pioneer work on computational vision. According to Marr, the ultimate task of any computer vision system is essentially to “transform” an array of input numerical values into meaningful description that a human normally perceives. Figure 1.1(a) shows an image which consists of two texture regions. However, the texture regions are not obvious at all from the numerical values, where a small portion is shown in Figure 1.1(c).

While the true “transformation” employed by humans is not known, any algorithm for solving vision problems attempts to approximate the transformation based on different assumptions and constraints with respect to the problems to be solved. Existing approaches can thus be categorized according to the problems to be solved and their assumptions. Due to the complexities of the vision process, four problems are widely studied in computer vision relatively independently: edge detection, stereo
Figure 1.1: A texture image and the corresponding numerical arrays. (a) A texture image with size $128 \times 64$. (b) A small portion with size $40 \times 30$ of (a) centered at pixel $(64; 37)$, which is on the boundary between the two texture regions. (c) Numerical values of (b). To save space, the values are displayed in hexadecimal format.
matching and motion analysis, image segmentation and perceptual organization, and pattern recognition. Roughly speaking, the techniques for the first three problems are primarily data-driven, or called bottom-up processes, and pattern recognition is model-driven, or called top-down process.

Early techniques with successful applications are classification techniques [25], which map a given input into one of the pre-defined classes according to a distance measure. However, all the possible classes and their variations we normally perceive are too gigantic to be implemented effectively in any system. The attention of computer vision was then shifted to derive more generic features for arbitrary images. From information and encoding theories, edges, i.e., discontinuities in images, carry more information and exhibit nice properties such as invariance to luminance changes. Motivated by neurophysiological findings [53], many edge detection algorithms were proposed and studied. Segmentation techniques try to solve the same problem by segmenting an image into homogeneous regions, where edges and region contours can be obtained straightforwardly and more robustly. These approaches were claimed to be unified [92] through what is called Mumford-Shah Segmentation energy functional [94] (See Chapters 4 and 5). Common to these approaches, the images are assumed to be piece-wise smooth regions with additive Gaussian noise, resulting in efficient algorithms. To improve the performance for real images, multiple scales are generally needed and linear and nonlinear scale spaces are thus proposed and studied. Chapter 2 studies segmentation for range images. Chapter 3 studies a new nonlinear smoothing algorithm and addresses some of problems in nonlinear scale spaces. Chapter 7 applies a nonlinear smoothing algorithm to hydrographic object extraction from remotely sensed images.
While there are useful applications of edge detection and segmentation algorithms, the underlying assumption limits their successes in dealing with natural images. As shown in Figure 1.1, texture regions neither are piece-wise smooth nor can be modeled with additive Gaussian noise. Figure 1.2 demonstrated that a pure linear system is not sufficient for natural image modeling [63], where the spatial relationships among pixels are more prominent in characterizing the texture regions than individual pixels. Clearly piece-wise smooth regions with additive Gaussian noise are not sufficient and more sophisticated models are needed to deal with texture images.

Supported by neurophysiological and psychophysical experiments [12] [23], the early processes in the human vision system can be abstractly modeled by filtering with a set of frequency and orientation tuned filters. However, as demonstrated in Figure 1.2, purely linear filtering is not sufficient, nonlinearity beyond filtering must be incorporated [87]. Spectral histograms integrate the responses of a chosen bank of filters through marginal distributions [148] [149] [150] [147]. As demonstrated in Figure 1.2, spectral histograms are nonlinear. Chapters 4 and 5 apply spectral histograms to modeling [147], classification, and segmentation of texture as well as intensity images.

While edge detection and segmentation techniques are very fruitful, there are perceptual phenomena that cannot be explained by purely data-driven processes. Classical examples include virtual contours, which are widely studied by Gestaltists. The long-range order grouping is known as perceptual organization. Chapter 6 studies perceptual organization through temporal dynamics.

Because many of the meaningful objects cannot be characterized well using intensity values or even textures, such as a human face, the relationships among some
Figure 1.2: Demonstration of nonlinearity for texture images. (a) A regular texture image. (b) The image in (a) was circularly shifted left and downward for 2 pixels at each direction. (c) The pixel-by-pixel average of (a) and (b). The relative variance defined in (3.20) between (a) and (b) is 137, and between (a) and (c) is 69. The distance between the spectral histograms defined in Chapter 4 between (a) and (b) is 1.288 and between (a) and (c) is 38.5762.
primitives need to be modeled. This leads to the need of top-down processes such as recognition. Clearly, the four problems studied are sub-problems of vision process and the integration among them is critical for a complete vision system. The interaction between different modules is briefly discussed in Chapter 8.

1.2 Thesis Overview

As discussed above, we study vision problems at different organizational levels in this dissertation. In Chapter 2, we study the segmentation problem for range images. Depth is most important cue for visual perception and range image segmentation has a wide range of applications. We propose a feature vector consisting of surface normal, mean and Gaussian curvatures and a similarity measure for range images. We implemented a system based on oscillatory correlation using a LEGION (locally excitatory globally inhibitory oscillator network) network. Experimental results demonstrate that our system is capable of handling different kinds of surfaces. With the unique properties of a temporal dynamic system, our approach may lead to a real-time approach for range image segmentation.

In Chapter 3, we propose a new nonlinear smoothing algorithm by incorporating contextual information and geometrical constraints. Several nonlinear algorithms are derived as special cases of the proposed one. We have compared the temporal behavior and boundary detection results of several widely algorithms, including the proposed method. The proposed algorithm gives quantitatively good results and exhibits nice temporal behaviors such as quick convergence and robustness to noise.

In Chapter 4, we propose spectral histograms as a generic statistic feature for texture as well as intensity images. We demonstrate the properties of spectral histograms
using image synthesis, image classification, and content-based image retrieval. We also compare with several widely used statistic features for textures and show that the distribution of local features is critically important for classification while mean and variance in general are not sufficient. We also propose a model for texture discrimination, which matches the existing psychophysical data well.

Chapter 5 continues the work in Chapter 4. In Chapter 5, segmentation problem is studied extensively using spectral histograms. A new energy functional for segmentation is proposed by making explicit the homogeneity measures. An approximate algorithm is derived, implemented and studied under different assumptions. Satisfactory results have been obtained using natural texture images.

Chapter 6 studies the problem of perceptual organization and long-range grouping, which is one level beyond the segmentation. By using a boundary-pair representation, we propose a figure-ground segregation network. Gestalt-like grouping rules are incorporated by modulating the connection weights in the network. The network can explain many perceptual phenomena such as modal and amodal completion, shape composition and perceptual grouping using a fixed set of parameters.

Chapter 7 presents a computational framework for feature extraction from remote sensing images for map revision and geographic information extraction purposes. A multi-layer perceptron is used to learn the features to be extracted from examples. A locally coupled LEGION network is used to achieve accurate boundary localization. To increase the robustness of the system, a weight adaption method is used. Experimental results using DOQQ images show that our system can handle very large images efficiently and may have a wide range of applications.
Chapter 8 summarizes the contributions of the work presented in this dissertation and concludes this dissertation with discussions on the future work.