CHAPTER 8

CONCLUSIONS AND FUTURE WORK

8.1 Contributions of Dissertation

This dissertation has investigated computational issues at different levels of image organization and the major contributions are:

- We propose a new similarity for range images and implement a range image segmentation system using a LEGION network.
- We propose a contextual nonlinear smoothing algorithm and show that several widely used nonlinear smoothing algorithms are special cases. The proposed algorithm generates quantitatively better results and exhibits nice properties such as quick convergence.
- We propose the spectral histogram as a generic statistic feature for texture as well as intensity images.
- We study image classification using the spectral histogram. We show that mean and variance as statistical features are not sufficient and the distribution of features is critically important for classification and segmentation.

- We propose a new energy function for image segmentation which expresses explicitly the homogeneity criteria for segmentation. We implement an approximate deterministic algorithm for image segmentation.

- We propose a method which can detect homogeneous texture regions in an input image using the relationships between different scales.

- We propose a novel method for precise texture boundary localization utilizing the structures of textures.

- We propose a boundary-pair representation and figure-ground segregation network using temporal dynamics for perceptual organization.

- We propose a new computational framework for extracting features from remote sensing images by combining the advantages of the learning-by-example methods and locally coupled oscillator networks for better boundary accuracy.

8.2 Future Work

8.2.1 Correspondence Through Spectral Histograms

Two major areas in computer vision that are not addressed in this dissertation are stereo matching and motion analysis. The central issue underlying both problems is how to establish correspondence between input images, known as the correspondence problem. We argue that the spectral histogram with the associated similarity measure would potentially provide a solution to the correspondence problem. Because the spectral histogram implicitly encodes the structures through marginal distributions, it reduces matching ambiguities significantly compared to cross-correlation and other feature-based matching techniques. For example, Figure 8.1 (a) and (b) show a stereo image pair of a bridge. Without any assumptions of camera positions and matching models, we extract the spectral histogram at a given pixel and find the matches in the paired image through search. Figure 8.1 (c)-(e) show three examples. where the middle image shows the probability of pixels in the paired image being a match of the given pixel. In all the three cases, the matched regions are identified uniquely and correctly. In Figure 8.1(e), the matched region is not localized because the pixels in the surrounding area are structurally similar. If we modify the algorithm for automatic homogeneous texture region extraction proposed in Chapter 5, we can essentially identify good features in one image and then find the matching region(s) in the paired image. Parameters for transformation between the images can then be estimated. From this example, one can see the correspondence problem can be potentially solved more effectively.

8.2.2 Integration of Bottom-up and Top-down Approaches

The purpose of a vision system is to localize and recognize important objects. To achieve that, different cues need to be integrated together. For example, contours have long been realized as an important feature in characterizing objects. However, obtaining reliable contours from natural images remains difficult; edge detection algorithms often give "meaningless" edges. While the spectral histogram incorporates only photometric properties of surfaces and objects, i.e., intensity values, meaningful contours can be extracted from the segmentation results using the algorithms proposed in Chapter 5. Figure 8.2(a) shows a natural image of a giraffe. Figure 8.2(b) shows a typical output from a Canny edge detector [13]. It is evident that deriving

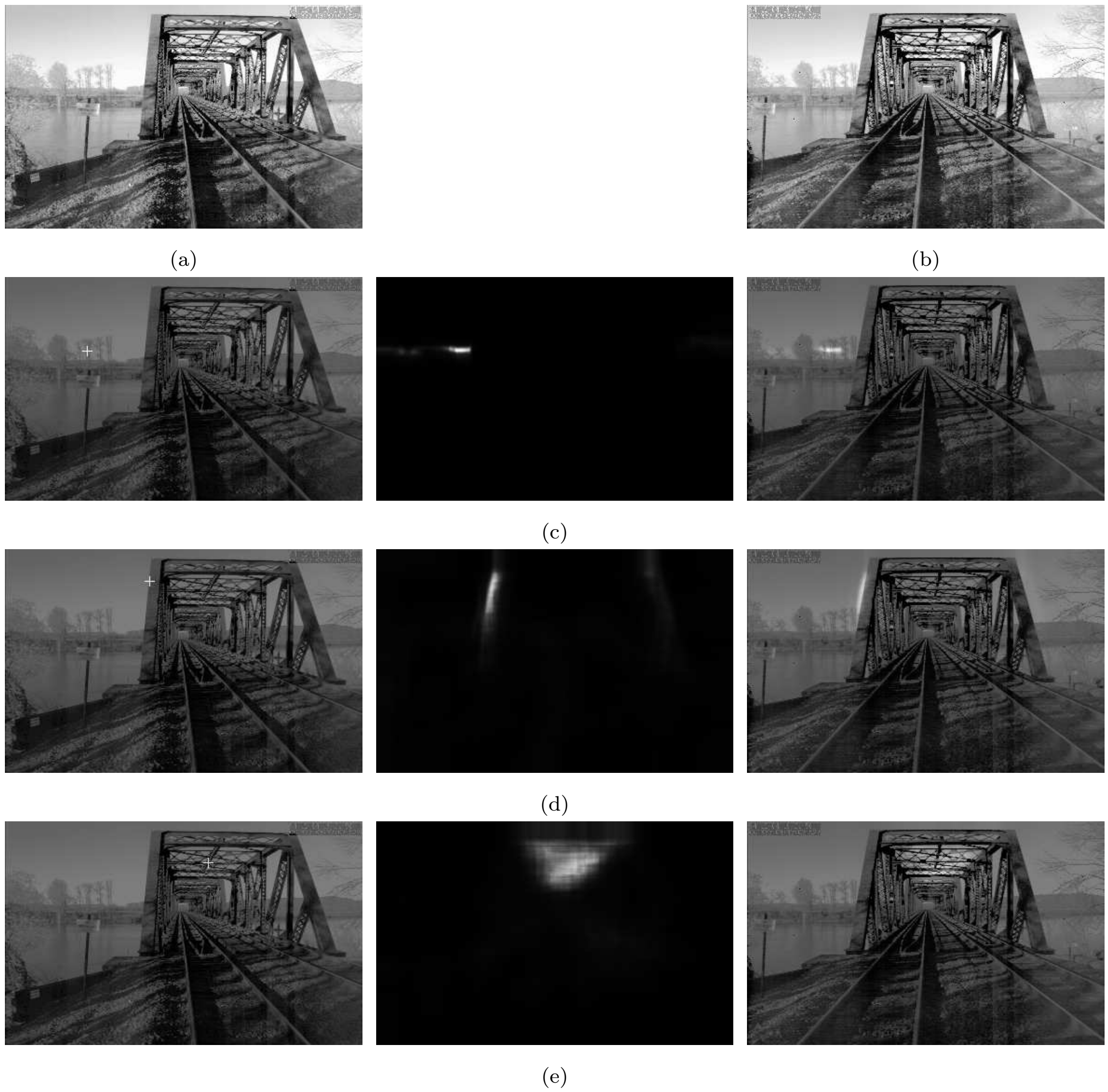

Figure 8.1: A stereo image pair and correspondence using the spectral histogram. (a) The left image. (b) The right image. (c)-(e) The matching results of marked pixels in the left image. In each row, the left shows the marked pixel, the middle shows the probability of being a match in the paired image, and the right shows the high probability area in the paired image.

a reliable contour from the generated edges is not feasible due to the local ambiguities. Figure 8.2(c) shows an initial segmentation result using the method proposed in Chapter 5. One can see that a contour of the giraffe can be obtained easily. This demonstrates that features from bottom-up processes must be generic.

Of course, no one can expect a perfect segmentation and recognition result from purely bottom-up algorithms. The top-down influence from recognition plays an important role in achieving the purpose of a vision system. For example, an iterative procedure can be initiated based on the result shown in Figure 8.2(c). In this case, both the photometric properties and the contour may suggest a giraffe with high probabilities. With the top-down knowledge, the segmentation result and contour can be improved by recovering the missing parts of the giraffe.

There are important computational issues that need to be addressed in order to model the interactions between bottom-up and top-down processes. In this regard, temporal correlation with LEGION as a concrete implementation provides an elegant representational framework. By utilizing temporal domain, LEGION provides distinctive advantages that are unique to dynamic systems and is biologically plausible. The current model of LEGION [123] [134] [135] employs only very local couplings, which significantly limits its potential. By incorporating longer-range and top-down couplings, a complete vision system is conceivable. We have obtained very promising results by integrating bottom-up and top-down processes [133], which is not included in this dissertation.

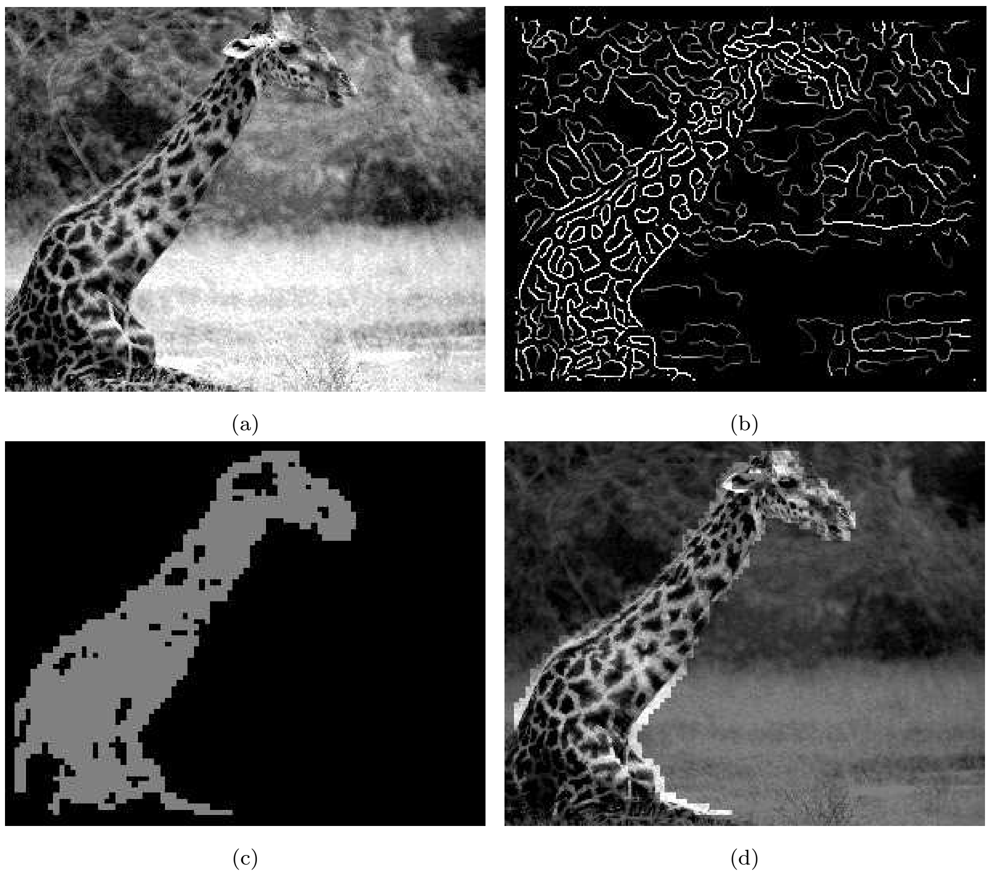

Figure 8.2: Comparison between en edge detector and the spectral histogram using a natural image of a giraffe. (a) The input image with size 300×240. (b) The edge map from a Canny edge detector [13]. (c) The initial classification result using the method presented in Chapter 5. A spectral histogram is extracted at pixel $(209, 291)$ and the segmentation scale is 29×29. (d) The initial classification is embedded in the input image to show the boundaries.

8.2.3 Psychophysical Experiments

While we claim that spectral histograms provide a generic feature for images and have demonstrated that by synthesizing a wide range of texture images, including regular patterns, texton images, and many other natural images, rigorous psychophysical experiments are needed to solidify the hypothesis. The result for texture discrimination discussed in Section 4.7 is very promising. However, the images used in the experiment are synthetic and thus are not representative for natural images. One straightforward experiment is to test on the correspondence between the observed and synthesized images, as shown in Figure 4.16 and Figure 4.13. Another experiment is to order a set of textures by humans as well as by an algorithm based on spectral histograms. Other experiments can utilize the texture synthesis tool we have to control the sharpness of texture boundaries and test on boundary accuracy and asymmetry in texture perception.

Given all the results we have achieved using spectral histograms, some of which match the human performance well, we would like to investigate if spectral histograms are biologically plausible. It is well known that neurons encode information through temporal spikes and spectral histograms can also be encoded very effectively that way because spectral histograms can naturally be represented using temporal spikes. Also the distance measure we used between two spectral histograms can be approximated through cross correlations between them. As hypothesized by von der Malsburg [130], temporal correlation provides a mechanism for solving fundamental problems in perception such as feature binding. Spectral histograms would then extend significantly the functionalities that can be achieved through temporal correlation.

8.3 Concluding Remarks

While developing a generic computational system for seeing remains the dream of many vision researchers, significant progress can certainly be made by pursuing the fundamental problems in a natural environment. Certainly there are many plausible approaches for computer vision and the criterion to compare them is how efficient vision tasks can be solved. Given the successful methods we have for relative independent problems such as segmentation and pattern recognition, next steps would be how to model the interactions among different modules and integrate them effectively into a complete vision system. It is my sincere hope that this work would provide some useful insights for solving the computer vision problems.