Preface

Computer vision has been a rapidly developing research area since the mid of 1970s. It is the science and technology of machines that see, and focuses on the understanding of digital input images, in many forms including videos and 3D range data. Although it is a diverse and relatively new field of study, it has achieved many scientific breakthroughs and has been successfully applied to a variety of challenging real-world applications. These include automatic face recognition, medical image analysis, and car registration plate recognition.

The process of computer vision can be divided into the stages of a) low-level processing, b) image feature extraction, and c) high-level visual understanding. At low-level, the focus is on how to extract geometrically invariant features in a robust manner so as to characterize the structures present in the visual data. Early contributions include the Harris corner detector and the Canny edge detector, which have both proved effective and have been used as the basis of higher level shape and object recognition. More recently, more robust feature extraction approaches have been proposed and these include SIFT (scale-invariant feature transform), Shape Contexts, and LSS (local self-similarity). These methods can be used to extract translation, scale, and rotation invariant features from local image patches. After low-level processing and feature extraction have been performed, high-level vision processing borrows ideas from both statistical and structure based machine learning to make final decisions concerning object classification or recognition and parameter estimation.

In computer vision, information concerning scene and object structure plays an important role. For example, after extracting SIFT features, it is not only the extracted feature vectors, but also the relative positions of these feature points that are crucial in the subsequent recognition stage. Structural pattern recognition has been a mature field of research for over three decades, and has played a pivotal role in computer vision. Importing ideas from the mathematics of graph theory is an effective and natural way to represent the structural information residing in a scene. There are also many well-developed algorithms for the analysis of graphs. Hence, graph based methods have found widespread use to solve problems in computer vision. Although there is an active and vibrant literature in the area, the use of graph-based methods for computer vision applications is a niche topic. This book aims to address problems related to applying graph-based methods in computer vision. The book includes accounts of the latest developments in graph-based methodology and its application to a variety of problems in computer vision. The list includes successful examples in the areas of image segmentation, image matching, and classification, where graph-based methods play a vital role. The remainder of the book is organized as follows:

Section 1 focuses on graph-based image matching. Image matching is a fundamental issue in computer vision. It is a vital step in object detection, stereovision, and image recognition. The basic idea is to find the correspondence parts between two images. Early attempts try to extract feature points i.e. using the
Harris or SIFT methods, in given images and then find the most similar pairs by computing the feature vector difference. Sophisticated statistical methods have been proposed recently i.e. RANSAC, to improve the consistency of the matching stage. However, the process of image matching should rely not only on the value of the extracted feature vectors, but also upon the positions of these feature points. The positions of the feature points can be used to construct a graph structure representing the arrangement of points, and then graph based methods can be used for image matching. In the graph-based literature, the methods for finding correspondences between nodes in different graphs are referred to as “graph matching.” However, graph matching is normally an NP hard problem. However, this problem is alleviated using optimization techniques such as mathematical relaxation. Many statistical methods have also been used to solve relaxed graph matching problems and these include the EM (Expectation and Maximization) algorithm and RANSAC. These methods normally iterate to the optimal solution. The iteration time depends on the specific algorithms used and the initial parameters chosen. One interesting research topic is how to combine probabilistic relaxation methods with machine learning. Image matching with learning can increase the performance, especially in situations where prior knowledge is available.

This section contains three chapters. The chapter “Graph Matching Techniques for Computer Vision” introduces and reviews the many different graph matching techniques that have been used for computer vision, and relates each application with the techniques that are most suited to it. The second chapter, “Geometric-Edge Random Graph Model for Image Representation,” casts image matching into a G-E random graph matching problem by using the random dot product algorithm. The third chapter, “The Node-to-Node Graph Matching Algorithm Schema” proposes a new graph matching algorithm which can be used for directed attributed graph matching and has been applied to the problem of scenario matching.

Section 2 discusses graph-based methods for image segmentation. Image segmentation is a mid-level process in vision processing. The basic idea is to segment the image into meaningful subparts. As mentioned earlier, in graphs can represent the structural information present in an image. The region adjacency graph produced by image segmentation provides a simple and effective representation of scene structure. Graph cut methods are used in graph theory to recursively bipartition a graph. The bipartition is effects by identifying edges which, when removed, give edge isolated cliques. Well known examples of graph cut algorithms include the normalized cuts method of Shi and Malik. This is a graph spectral method which uses both the total dissimilarity between the different subgraphs together with the total similarity within them. Subsequently, many improvements have been proposed to normalized cuts. Recent examples include, hierarchical segmentation, which can better reflect perceptual models of image segmentation. Graph methods play a critical role in hierarchical image segmentation, where they capture not only structural relationships at the same level of representation but also in the relationships between different levels. Prior knowledge can also be used within the segmentation stage, especially to capture object-part sub-structure. The required prior knowledge can be extracted from training samples using machine learning techniques. By representing object-part relationships using a graph, the image segmentation process can incorporate semantic knowledge. These methods can additionally be used in object detection.

In Section 2 of the book, the editors and authors present four chapters concerning graph-based methods for image segmentation. The chapter “Unsupervised and Supervised Image Segmentation Using Graph Partitioning” presents a method for image segmentation using an improved graph partition algorithm. The integrated probabilistic models in this paper increase the segmentation performance compared with that obtained using traditional graph partition segmentation algorithms. The chapter “Motion Segmentation and Matting by Graph Cut” uses graph cuts to implement motion segmentation, and is then applied graph
to motion analysis. The chapter “Hypergraph Based Visual Segmentation and Retrieval” imports ideas from hypergraph analysis for segmentation. “Recent Advances on Graph-Based Image Segmentation Techniques” gives a review of the available graph-based methods for image segmentation.

Section 3 concerns image or video classification based on graph-methods. Image recognition is one of the most important tasks in computer vision, pattern recognition, and machine learning. Since the inception of the field, considerable attention has been devoted to image and object recognition. Early attempts focused on simple objects occurring in face images, fingerprint images, and iris images. The images studied were normally devoid of complex background structure. Traditional methods such as histogram-based classification cannot incorporate the spatial arrangement of the objects. For example, in face recognition, graph structure can be used to represent the arrangement of facial features. The use of spatial arrangement information can increase the recognition performance. Moreover, when used in conjunction with problems such as face detection and face recognition, spatial arrangement information leads to algorithms that can function effectively even when there is a complex background. Pictorial structure is another example where such benefits can be reaped using graph structure. For instance deformable structure model can be used for object detection. Here the deformable model can capture high-level knowledge that can be used for scene understanding based on the arrangement of objects.

Graph representations are an effective way to represent pictorial structure and one particularly effective class of methods are based on spectral graph theory. When an object is represented by a graph structure, graph spectral methods can be used to extract a fixed length feature vector to characterize the structure of the object. Then tradition pattern classification techniques can be used for recognition. Central to the successful use of graph spectral methods in image recognition is the issue of how to efficiently extract stable and invariant features from the eigenvalues and eigenvectors of a suitable matrix representation of the graph. Many algorithms have been proposed and an evaluation of the different alternatives is therefore necessary. For complex images with background, more complex algorithms are needed. Image segmentation can be combined with graph-based methods for recognition. Another application for graph-based methods is shape analysis. In shape analysis, shapes can be represented by graph structures. The analysis on shapes then transferred to graph structure analysis.

Section 3 includes seven chapters on image and video recognition or classification. “Graph Embedding Using Dissimilarities with Applications in Classification” uses graph embedding to find a similarity measure in the embedded space and then uses this to classify graphs. The chapter “Generative Group Activity Analysis with a Quaternion Descriptor” uses graphical models to analyse group activities. “Shape Retrieval and Classification Based on Geodesic Paths in Skeleton Graphs” extends traditional shock graph methods to better characterize shapes for classification and retrieval. The chapter “Discriminative Feature Selection in Image Classification and Retrieval” describes a new method for discriminative feature selection which can be used to improve the performance of image classification and retrieval. “Normalized Projection and Graph Embedding via Angular Decomposition” introduces graph embedding for image analysis. “Region-based Graph Learning towards Large Scale Image Annotation” concerns image annotation using graph-based methods. Finally, in “Copy Detection Using Graphical Model: HMM for Frame Fusion,” the authors use graphical model for video analysis for copy detection.

Finally, in Section 4, there are two chapters that use graph based methods to solve image-processing problems. The chapter “Multi-Scale Exemplary Based Super-Resolution with Graph Generalization” introduces a coding system with a resolution-invariance property, such that it is able to handle continuous-scale image resizing. This is to be compared with traditional methods that only support single integer-scale upsizing. This chapter generalizes the graphical model to the case where the typical non-linear
coding process is approximated by an easier-to-compute function. Thus, the SR process can be highly parallelized by modern computer hardware. As demonstrated by the chapter, the proposed system gives very promising image SR results in various aspects. In the chapter “Graph Heat Kernel Based Image Smoothing,” heat kernel diffusion process is used to smooth the images. Heat kernel relates to graph spectral theory and contains structural information of the graphs. Image smoothing is accomplished by convolving the heat kernel with the image, and its numerical implementation is realized by using the Krylov subspace technique. The method has the effect of smoothing within regions, but does not blur region boundaries.

To conclude, although they have found a useful niche, the widespread use of graph-based methods to solve traditional computer vision problems has yet to take place. This is a disappointing and represents a missed opportunity. Almost all problems in computer vision are formulated over arrangements of objects. Although techniques such as projective geometry capture the details of projection from scene to camera plane, they do not furnish a natural way of representing the types of relational or semantic detail necessary for scene or image understanding. Moreover, even at low level, graph representations are a more natural way to capture image models expressed over discrete pixel lattice, and avoid problems involved in discretising intrinsically continuous models such as those furnished by partial differential equations. This said, there is also a research imperative to explore further and more demanding computer vision application areas to understand how to apply graph-based algorithms to a greater variety of problems. After decades of development, a rich literature exists in this area. In this book, the contributors present a flavour of the research issues related to applying graph-based methods in computer vision. This book will be both timely and of value to the research community. It not only provides a snapshot of current activity in the field, but also provides a reference work that, it is hoped, will attract others to this fascinating topic, and hopefully inspire them to develop both novel graph-based method in more challenging application area, which in turn will lead to vision systems with better performance.