Chapter 1

INTRODUCTION

In recent years, with the advent of newer, much improved and inexpensive imaging technologies and the rapid expanding of the Internet, more and more images are becoming available. Recent developments in image collection platforms produce far more imagery than the declining ranks of image analysts are capable of handling due to human work load limitations. Relying on human image experts to perform image analysis, processing and classification becomes more and more unrealistic. Building object detection and recognition systems to take advantage of the speed of computer is a viable and important solution to the increasing need of processing a large quantity of images efficiently.

1.1 Object Detection and Recognition Problem

The object detection and recognition problem is one of the most important research areas in pattern recognition and computer vision [7], [18]. It has wide range of applications in surveillance, reconnaissance, object and target recognition, autonomous navigation, remote sensing, manufacturing automation, etc. The major task of object detection is to locate and extract regions that may contain objects in an image. It is an important intermediate step to object recognition. The extracted regions are called regions-of-interest (ROIs) or object chips. ROI extraction is very important to object recognition,

since the size of an image is usually large, leading to the heavy computational burden of processing the whole image. By extracting ROIs, the computational cost of object recognition is greatly reduced, thus improving the recognition efficiency. This advantage is particularly useful to real-time applications, where the recognition speed is of prime importance. Also, by extracting ROIs, the recognition system can focus on the extracted regions that may contain potential objects and this can be very helpful in improving the recognition accuracy. Generally, the extracted ROIs are identical to their corresponding regions in the original image, but sometimes, they may be images that result from applying some image processing operations to the corresponding regions in the original image. No matter what ROIs are, they are passed to an object recognition module for further processing. Usually, in order to increase the probability of object detection, some false alarm ROIs, which do not contain an object, but some natural or man-made clutter, are allowed to pass object detection phase.

The task of object recognition is first to reject the false alarm ROIs and then recognize the kinds of objects in the ROIs containing them. It is actually a signal-to-symbol problem of labeling perceived signals with one or more symbols. A solution to this problem takes images or the features extracted from images as input and outputs one or more symbols which are the labels of the objects in the images. Sometimes, the symbols may further represent the pose of the objects or the relations between different objects. These symbols are intended to capture some useful aspects of the input and in turn, permit some high level reasoning on the perceived signals.

It is well known that automatic object detection and recognition is really not an easy task. The quality of detection and recognition is heavily dependent on the kind and quality of features extracted from the image, and it also highly relies on the representation of an object based on the extracted features. The features used to represent an object are the key to object detection and recognition. If useful features with good quality are unavailable to build an efficient representation of an object, good detection and recognition results cannot be achieved no matter what detection and recognition algorithms are used. However, in most real images, there is always some noise, making the extraction of features difficult. More importantly, since there are many kinds of features that can be extracted, so what are the appropriate features for the current detection and recognition task or how to synthesize composite features