Chapter 4

FEATURE SELECTION FOR OBJECT DETECTION

4.1 Introduction

The goal of feature selection is to find the subset of features that produces the best object detection and recognition performance and requires the least computational effort. Feature selection is important to object detection and recognition systems mainly for three reasons:

First, using more features can increase system complexity, yet it may not always lead to higher detection/recognition accuracy. Sometimes, many features are available to a detection/recognition system. However, these features are not independent and may be correlated. A bad feature may greatly degrade the performance of the system. Thus, selecting a subset of good features is important.

Second, features are selected by a learning algorithm during the training phase. The selected features are used as a model to describe the training data. Selecting many features means a complicated model is used to approximate the training data. According to the minimum description length (MDL) principle, a simple model is better than a complex model [100]. Since the training data may be corrupted with a variety of noises, a complex model may

overfit the training data. Thus, a complex model may be sensitive to noise in the training data and its performance on unseen test data may be bad.

Third, using fewer features can reduce the computational cost, which is important for real-time applications. Also it may lead to better classification accuracy due to the finite sample size effect.

In this chapter, we use genetic algorithm (GA) [41], [42], [77] to select as few features as possible to describe the training data effectively. The specific application we focus on is the detection of targets in SAR images. Automatic detection of potential targets in SAR imagery is an important problem [7], [65]. A constant false alarm rate (CFAR) detector is commonly used to "prescreen" the image to localize the possible targets [65]. Generally, targets correspond to bright spots caused by strong radar return from natural or man-made objects. Parts of the imagery that are not selected are rejected from further consideration. In the next stage of processing, regions of interest are further examined to distinguish man-made objects from natural clutter. Finally, a classifier such as a Bayesian classifier, a template matcher or a model-based recognizer is used to reject man-made clutter.

GAs are widely used in image processing, pattern recognition and computer vision [7], [13], [23]. They are used to evolve morphological probes that sample the multi-resolution images [101], to generate image filters for target detection [53], to select good parameters of partial shape matching for occluded object recognition [89], to perform pattern clustering and classification [113], etc. GAs are also used to automatically determine the relative importance of many different features and to select a good subset of features available to the system [95].

The problem we address is to select a minimal set of features to distinguish targets from natural clutter. The approach is based on a closed loop system involving GA based feature selection and a Bayesian classifier. GA uses a MDL-based fitness function that combines the number of features to be used and the error rate of the classifier. The results are presented using real SAR images. The experimental results show that the MDL-based fitness function is the most effective in selecting a minimal set of features to describe the data accurately compared to other three fitness functions, and the subset of features