MORPHOLOGICAL IMAGE PYRAMIDS

FOR AUTOMATIC TARGET

RECOGNITION

By

WEI CHEN

Bachelor of Science
Nanjing University
Nanjing, China
1982

Master of Science
Shanghai Institute of Technical Physics
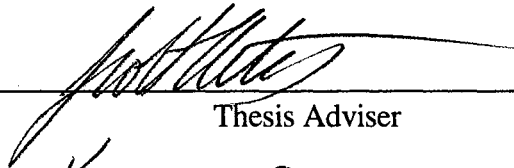Shanghai, China
1988

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
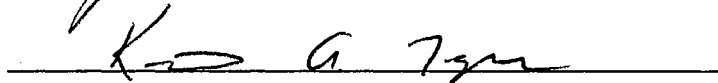July, 1997

OKLAHOMA STATE UNIVERSITY

Thesis
1997D
C5775m

MORPHOLOGICAL IMAGE PYRAMIDS

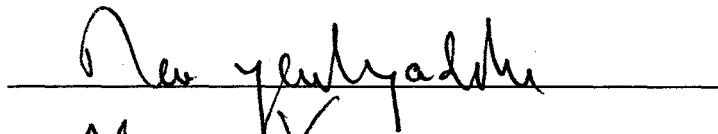# FOR AUTOMATIC TARGET

RECOGNITION

Thesis Approved:

_____
Thesis Adviser

_____

_____

_____

_____
Dean of the Graduate College

ACKNOWLEDGMENTS

First of all, I would like to express my great appreciation to Dr. Scott T. Acton, my thesis advisor, for his intelligent supervision and constructive guidance during the entire period of my dissertation research. Without his support and encouragement this thesis could not have been written. Secondly, I would also like to extend my special appreciation to Dr. Keith A. Teague, Chairman of my advisory committee, for his generous help, constant guidance, and friendship throughout the past few years. I am grateful to my other committee members, Drs. Rao Yarlagadda and Xin-cheng Xie, for their invaluable advice.

In the course of preparing and writing this thesis, many people have assisted me and offered their support. I have benefited significantly from C. Andrew Segall, Walter D. Andrews, Anthony S. Wright, Randy Bartles, and Brian D. Little, for their academic suggestions and technical assistance. I am further indebted to Dr. Wei-ping Li of Department of Mathematics, and Ms. Yihong Xu of RPI.

Finally, I want to give my special thanks to my husband, Wei Shan, for his love, sharing, understanding, strong encouragement, support, and various suggestions to my research. I would like to thank my parents for their love, constant encouragement and

invaluable support. I would also like to thank my son, David, for his patience and understanding throughout this long year. This thesis is dedicated to them all.

iv

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Over the past two decades, we have witnessed a tremendous development of image processing techniques. The field of image processing has grown considerably with the increased utilization of imagery in a myriad of applications coupled with the improvements in size, speed, and cost effectiveness of computers and related signal processing technologies. Image processing has evolved from academic interest into significant roles in scientific, industrial, biomedical, space, and government applications.

Image processing is a broad subject encompassing studies of physics, physiology, electrical engineering, computer science, and mathematics. It involves a variety of technical disciplines and has numerous practical applications. This thesis is devoted to studying one of the challenging frontiers of image processing: *Automatic Target Recognition* (ATR).

ATR refers to the identification of potential targets, and the recognition of actual targets without humans in the process. This is desirable since a system with a human in the loop is generally slow, unreliable, vulnerable, and may limit the performance of the overall system (Walker, 1983). An important application of ATR is in helping and guiding pilots

of high-performance aircraft flying close to the ground during inclement weather or at night.

Basically, an ATR system performs automatic target acquisition, identification, and tracking by processing a sequence of images. A wide variety of ATR algorithms have been reported in the literature (Bhanu, 1986). A conventional ATR system incorporates time-consuming preprocessing, segmentation, correlation, and classification operations on each image in a video sequence at its original spatial resolution. This traditional fixed resolution ATR system spends more computational time on irrelevant image details, and requires a long time to identify an object. Such an approach is not feasible for a real-time implementation. In this thesis, an investigation of expeditious morphological pyramid identification and tracking algorithm which aims to improve the computational efficiency and solution quality of the traditional ATR system will be presented. This novel approach utilizes the nonlinear morphological filters integrated with multi-resolution pyramid scheme, which is motivated by the human visual system (Burt, 1989).

An image pyramid represents images in multi-resolution by a sequence of images created by successively filtering and subsampling the original image. Filtering is an image operation which can be either linear or nonlinear. With linear filtering, the pixel at coordinates (i,j) in the output image is a linear combination of the image intensities around (i,j) in the input image. For example, $O(i,j) = \frac{1}{3}[I(i,j) + I(i+1,j) + I(i-1,j)]$, where O denotes the output image intensity and I denotes the input image intensity. In contrast, a nonlinear filter implements a nonlinear function of neighboring image intensities. The nonlinear filters cannot be realized by linear convolution and have no frequency domain

characterization. Generally, the subsampling is carried out by discarding a number of rows and columns in the input image.

Early approaches in computer vision used linear lowpass filters, e.g., convolutions with a Gaussian-shaped kernel, to generate multi-resolution pyramids (Burt, 1989). This approach suffers from its shifting and blurring of important image features such as edges. In contrast, nonlinear morphological filter can preserve edges and the outline of object shapes at various feature scales. In this research, nonlinear morphological filters are utilized to generate multi-resolution pyramids for ATR.

The morphological filter is based on mathematical morphology which was developed in 1964 at the Paris School of Mines, France, by G. Matheron and J. Serra (Serra, 1982). Mathematical morphology is a set-theoretic method for image analysis. It provides a quantitative description of the geometric structures of an image. In mathematical morphology, a black and white (binary) image is represented by 2 dimensional Euclidean sets. If the background is black and the objects are white, the set of all white pixels forms a complete description of the binary image. The elements of the set are the coordinates of the white pixels in the image. Morphological operations use a pre-defined shape (set), called structuring element (SE), to interact with the image set. Therefore, each morphological operation is based on shapes. Since shape is a prime carrier of information in images, the importance of mathematical morphology is self-evident. By carefully selecting the SE, which is comparable to moving windows in typical nonlinear filters, morphological operations can simplify image data, preserve their essential

shape characteristics, and eliminate irrelevancies. They are therefore well suited for shape estimation which will lead to effective ATR.

The conventional ATR system based on a single resolution image is computationally inefficient. In contrast, the proposed ATR algorithm is based on multi-resolution --- the resolutions of the images in the pyramid vary from fine (original image) to coarse through successive subsampling. Therefore, it provides a hierarchical computing paradigm which may be utilized to decrease the computational cost of the automatic target recognition and tracking tasks. The pyramid allows the initial detection of objects at low spatial resolution and refinement of the final verification of the results at progressively higher resolution, but within smaller regions of the scene. In this way, computation time will be reduced by organizing search from coarse to fine. Burt has shown that when used in search application, these pyramidal structures can have a significant computational advantage (Burt, 1981). Thus, the multi-resolution representation provides a powerful tool for ATR by locating potential objects of interest within a scene using the most appropriate resolution instead of the fixed resolution in the traditional ATR algorithms.

In this research, the multi-resolution pyramid is combined with a nonlinear morphological filter to provide an effective tool for ATR. The morphological pyramid (MP) creates a multi-resolution representation through a sequence of filtered images obtained by morphological filtering and regularly subsampling.

The morphological filter used in the MP will remove noise effectively and preserve the crucial features in an image at the same time. The edge and shape preservation properties of the morphological pyramids should provide improvements over detection

and recognition in a linear pyramid paradigm such as the Gaussian pyramid. The simplicity and computational efficiency provided by the MP make it appealing for real-time implementation. This will lead to enhancing the performance of the ATR system.

Given below, is an overview of the automatic target recognition scheme that exploits the structure of the nonlinear morphological pyramids utilized in this research. The scheme involves the generation of the MP, location and extraction of a region of interest (ROI), segmentation of the image, performance of the coarse-to-fine template matching to detect the target and tracking.

The first step is to develop a nonlinear pyramidal scheme that utilizes shape-preserving morphological operations to remove noise and unwanted detail. The optimal morphological filter and subsampling scheme used in generating MP needs to be studied. Since the performance of morphological operations is highly dependent on the size and shape of the structuring element, the SE used in morphological operations is investigated. The goal in subsampling is to preserve the integrity of the target information in each pyramid level and eliminate irrelevant details.

The choice of the pyramid level used for initial identification of targets must be analyzed. The determination of the initial level is a critical selection since the instance of a target missed at this level will never be recovered. In general, the selection depends on the original spatial resolution, the target size, the minimum distance between objects and the maximum shift in position allowed to the target in the search space, etc. Here, the morphological filtering at coarser levels can be quantified by region homotopy (the connection between several parts of the objects). The homotopy is important because it

provides information about the structure of an image. When the image is subsampled, the homotopy must be preserved. The evaluation of the homotopy for the morphological pyramid (MP) will prevent region merging and allow target/background discrimination for coarse (sparsely sampled) representations.

When a proper SE size and initial pyramid level is determined, the noise and unwanted detail in the constructed MP will be reduced, the uneven background will be smoothed, and the shape of the target will be preserved. Therefore, the MP is an effective engine for the ATR.

The initial step in the ATR process is to locate the region where a candidate target is present. By utilizing the MP structure, the target area can be easily extracted. An adaptive thresholding technique is applied to segment the potential target at the coarser level providing the approximate location of the candidate target. Once the result of this first screening of the region of interest is obtained, the second phase of refinement has to be performed. The refinement is accomplished step by step at every intermediate level until the maximum resolution level (the original image) is reached. At each level the confidence of the selection is checked, and the spatial position of the target is refined by checking the g nearest neighbors corresponding to the position of the previous level.

After the location of the ROI is determined, the ROI in the original image is extracted. Using this ROI, the edge-based template morphological pyramid is generated. Then, a coarse-to-fine template recognition scheme that exploits the structure of the morphological image pyramids is performed. In this pyramid-based target identification, potential targets are initially located by searching at lowest resolution. In this case the

search is formulated as the correlation of a low resolution template with a low resolution level of the image pyramid. Once a candidate target is found, the search procedure moves to higher resolution pyramid level and examines details that should occur within the target. The instance of the target located by this pyramidal template matching algorithm is input to a Kalman filter which predicts the location of the target in the next frame.

This pyramidal tracker can continue to track the target until target scale, target rotation, or change of viewpoint cause the matching between template and image fall below a threshold. At this point, a new template can be generated using the ROI guided by a Kalman predictor.

The remainder of this thesis is organized as follows. In Chapter II, an overview of the traditional automatic target recognition (ATR) algorithm, which includes preprocessing, detection, segmentation, classification, and tracking is described. Then the fundamental concepts of morphological image processing and the multi-resolution representations (pyramids) are briefly summarized. Chapter III discusses the construction of the optimal MP for ATR. By utilizing the morphological sampling theorem and homotopy preserving critical sampling theorem, the optimal morphological filter, the structuring element, and the initial identification pyramid level for identification are theoretically analyzed. The relationship between morphological sampling and the effect of the sampling on the homotopy is also presented. In Chapter IV, the proposed MP-based automatic target identification and tracking algorithms are detailed. Chapter V applies MP-based automatic target identification and tracking algorithms to infrared (IR) image

sequences. The comparative results are presented. Finally, the conclusions and

recommendations for future research are summarized in Chapter VI.

CHAPTER II

BACKGROUND

In this chapter, an overview of existing automatic target recognition algorithms is presented. The limitations of these ATR techniques are indicated. To overcome the disadvantage of these traditional ATR algorithms, the multiresolution morphological pyramid based ATR technique is proposed. In order to give the background of the morphological pyramid, it is necessary to briefly discuss the history of the mathematicalal morphology, the application of the mathematicalal morphology to image processing, and the multiresolution representations.

Automatic Target Recognition (ATR)

There are a large number of ATR algorithms. Typical ATR algorithms consist of several steps: preprocessing, detection, segmentation, feature computation, selection and classification, and tracking, as shown in Fig. 2.1 (Bhanu, 1986). The major disadvantage associated with these methods is the amount of computation needed to process the imagery at the original resolution. Brief descriptions of the standard ATR components are given below.
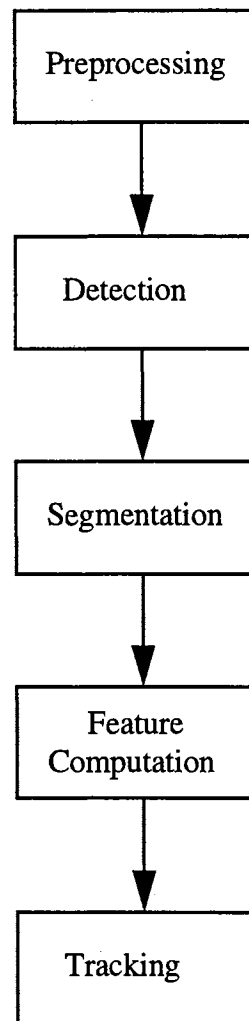
Fig. 2.1 Basic block diagram of typical ATR system

Preprocessing: This step is executed to improve target contrast and reduce noise and clutter present in the image. It is usually accomplished by a local filter such as the median filter for noise suppression, high-pass filters for edge sharpening, and locally variable scaling for contrast stretching. The limitation of pre-processing is that the error introduced in this step will affect the later detection. In this research, there is no such preprocessing step. The morphological pyramid has the intrinsic capability to eliminate noise.

Target Detection: This is the process of identifying those regions in the image that are likely to contain targets. The most common way to extract targets from the picture is to use thresholding technique in which localization and segmentation are inseparable. Burton and Benning (1981) compared four target detection algorithms: the "superslice" developed by University of Maryland (Milgram and Rosenfeld, 1978), the Ford double-gated contrast filter (Politopoulos, 1980), the "Spoke Filter" developed by the Army Missile Command (Minor and Sklansky, 1981), and the "contrast box" developed by Texas Instruments (Burt and Benning, 1981). All these algorithms require different parameter choices and have a high computational cost.

The "superslice" algorithm (Milgram and Rosenfeld, 1978) first generates a thinned edge map. Simultaneously, it applies a multi-threshold to the original image in order to extract different objects, and then uses a connected components algorithm to locate the borders. It then compares the thinned edge map with the sequence of thresholded images. The segmentation threshold is selected when the best match between the thinned edge and the border of the target is found.

The double-gated filter (Politopoulos, 1980) uses two rectangular windows. An inner window contains the target, and an outer window contains the immediate background of the target window. The filter localizes potential target areas based on the contrast between the target and its surrounding background.

The spoke filter (Minor and Sklansky, 1981) is an extension of the Hough circle detector (Pratt, 1991). It detects the object edges first and forms a spoke of a certain length and orientation perpendicular to each local edge.

The contrast box (Burt *et al.*, 1981), like the double-gated filter, uses a double window filter to locate candidate target areas. It detects objects based on the statistical difference between the two regions. The main difference between the double-gated filter and the contrast box lies in the matrices used to determine the likelihood of a target being localized at a certain pixel.

All the detection algorithms mentioned above process images in the fixed resolution (original image) and are time consuming. The detection method used in this research processes images in multiresolution. It first detects the region of interest (ROI) on a coarse resolution, for instance, 32 by 32 instead of 256 by 256 in traditional detection methods. Then, the target is detected in finer resolution, but only in the small regions guided by the ROI. Therefore, it is computationally effective compared to the traditional detection algorithms.

Segmentation: Once a potential target is localized, the segmentation step extracts the target from the background. Bhanu (1986) summarizes many segmentation algorithms for ATR which include superslice, pyramid approaches and relaxation. The pyramid spot

detection approach (Schneier, 1983) for extracting compact objects from a contrasting background involves constructing a succession of lower resolution images (a pyramid) and detecting spots in these images. Spots are detected by comparing each pixel with its 8 nearest neighbors. To obtain segmentation, thresholds can be calculated in the low resolution image or the original image. A variation of the above technique is the pyramid linking approach (Burt *et al.*, 1981) based on the creation of links between pixels in the successive levels of a pyramid. Segmentation is implemented as the process which selects a single legitimate parent for each node from that node's four candidate parents. The legitimate parent is the candidate with a value most like that of the node itself. Due to this nearest-parent selection rule, the segmented objects tend to be homogeneous regions, the objects boundaries are sharp, and noise can be smoothed out.

The above mentioned pyramid segmentation techniques use linear filters to construct pyramids which cause the object to merge with the background when the contrast between the object and background is small. In this research, a nonlinear morphological pyramid is applied to the pyramid linking technique to segment the target from the background (Chen and Acton, 1997). This is the first approach to utilize a morphological pyramid linking technique in segmentation. The segmentation and resulting edge detection yielded by this method is particularly effective in the presence of impulse noise. It demonstrates superior solution quality over standard full resolution detectors and previous pyramidal approaches.

Feature Computation, Selection, and Classification: After the targets are extracted, a set of features of them are computed for target classification. Most features

used by ATR researchers are either geometric, topological, or spectral. In this research, geometric features are extracted for identification.

Tracking: This is the process that predicts where the target is going to be in the next frame after finding the target for the first time. Once we have an estimate of the next target position, we can look at a small sub-image for the target and avoid searching the whole image.

Ayala *et al.* (1982) report a moving target tracking algorithm using a symbolic approach for image registration and motion analysis. Image registration is the process of locating corresponding points in two or more images of the same scene taken at different times and at different positions. The algorithm they used consists of several steps. The first step is segmentation using a region growing technique. The matching of corresponding segments between two images is based on the similarity of corresponding feature values. The features used in their matching include segment position, size, intensity, area ratio, and local contrast. Objects are detected as moving if they do not have a motion consistent with other objects in the image. This symbolic matching permits the matching of real objects which may have changed appearance slightly between images. Holben (1980) uses cross-correlation of successive images over subregions to estimate the scene shift so that the target can be tracked. When the confidence level of a tracker based only on intensity, feature, contrast, or correlation is low, tracking approach combined intensity, contrast, feature and correlation complements each other by allowing switching from one approach to another and results in a high confidence level. Dorrough *et al.* (1982) presents such a multimode tracker consisting of an intensity centroid tracker, an

edge tracker, and a correlation tracker. These tracking algorithms suffer from frequently lost targets in high clutter and low signal to noise ratio conditions (Bhanu, 1986). In addition, they are complicated and computationally intensive.

The Kalman filter is extensively used to compute a recursive solution to the problem of the least square estimation. It has been applied in numerous practical situations, including space navigation guidance, orbit determination, and tracking (Sage and Melsa, 1982). It is an optimal linear filter in the least mean squared error sense, and it can optimally estimate the target motion from noisy data. In this thesis, the Kalman filter is integrated with the MP to track the target.

In this section, the processing steps which constitute the typical approach to the ATR problem have been discussed. An overview of the techniques used to solve the ATR problem and the limitations of these techniques have been presented. The main disadvantage of the conventional ATR algorithms is that their computational load is quite high and their performance is not good in a low signal to noise ratio circumstance. The major difference between the traditional approaches and the approach used in this thesis is that the traditional algorithms process images in fixed resolution and the proposed method processes image in multiresolution to speed up and enhance the solution quality.

Every sensor is subject to noise. The conventional ATR systems use pre-processing to remove noise. As mentioned before, there is no preprocessing step in the proposed ATR algorithm. The noise is reduced by the morphological filter and subsampling operation in generating a morphological pyramid. Linear and nonlinear filters are used extensively in image processing to remove noise. It has been shown by many

experimental studies that although linear filters possess good noise attenuation capabilities, they smear the edges in the original image because of the linear average operation that they perform (Fong *et al.* 1989). On the other hand, nonlinear filtering is well-known as an edge-preserving method. This thesis will focus on target *detection*, since detection is the most vital step of the ATR system. Mathematicalal morphology offers an effective tool for target detection and noise reduction since it is a nonlinear processing operation which is the way that the human visual system works. Currently there is considerable research being performed on morphological image processing. The following section reviews the applications of morphological operations to image processing.

Morphological Operations

The word *morphology* refers to the study of form and structure. The morphological approach is generally based upon the analysis of a binary image in terms of some predetermined geometric shape known as a *structuring element* (SE). The initial theoretical work in mathematicalal morphology was done by Hadwiger (1957). Matheron (1975) developed it further in conjunction with integral geometry and size measurement, Kirsch (1957) first showed its utility in image processing, and Serra (1982) produced the first systematic theoretical treatment of the subject.

A morphological algebraic system is useful because compositions of its operators can be formed such that when acting on complex shapes, they are able to decompose them into their meaningful parts and separate them from their extraneous parts. Such a system

of operators and their compositions permits the underlying shapes to be identified and optimally reconstructed from their distorted, noisy forms. Furthermore, they permit each shape to be understood in terms of a decomposition, each entity of the decomposition being some suitably simple shape.

A familiar example of a non-morphological algebraic system is the algebraic system of convolution and its frequency domain representation. Here any finite duration function f can be represented as a sum of sinusoids. A distortion of f can be modeled by adding to it sinusoids which may not have been part of its original representation. Alternatively, f can be modeled by convolving the image with some kernel k. Whatever the distortion, understanding what happens in terms of the sinusoidal (frequency domain) representation permits one to develop procedures to remove the distortion of the undesired transformation. The underlying f can be estimated, reconstructed, extracted, or recognized on the basis of observing its distorted form. What the algebra of convolution does for frequency selection, the algebra of mathematicalal morphology does for shape.

Originally, mathematicalal morphology was developed for binary images. But from the very beginning there was a need for a more general theory including spaces of functions modeling gray-scale images.

Historically, morphological gray-scale image processing originally treated the gray functions as sets of layers of binary cross sections (thresholds). Processing was based on a stereological approach where cross sections are transformed individually or in pairs (Serra, 1982). Meyer (1977) developed gray-scale image contrast descriptors based on differences of openings of the functions by three dimensional structuring elements

composed of disks. Rosenfeld (1970) proposed a generalization of connectivity to functions, from which he derived a gray-scale thinning algorithm in collaboration with Dyer (1977). Goetcharian (1979) has also contributed several original notions such as the lower skeletons and relations of the fuzzy logical concepts to gray-scale morphological processing (Goetcharian, 1980). Serra (1982) has introduced formal notions of grayscale homotopy.

Morphological operations can be employed for many purposes, including edge detection, segmentation, and enhancement of images (Giardina and Dougherty, 1987; Chen and Acton, 1996). Schonfeld and Goutsias(1991) applied morphological filters to the problem of pattern restoration from noisy binary images. They proved that the class of alternating sequential filters is a set of parametric, smoothing morphological filters that "best" preserves the crucial structure of input images in the least mean difference sense. This filter "optimally" eliminates the rough characteristics of the degradation noise while it "optimally" preserves the crucial geometrical and topological features of the noise-free pattern.

Mathematicalal morphology has recently attracted a great deal of attention from the image processing world (Salembier *et al*, 1997, Maragos, 1997, Sidiropoulos *et al*, 1997, Morales, 1995), primarily due to the fact that it concerns the shape and properties of objects (regions) of an image and how these may be changed to extract useful features. From the underlying morphological operation, an entire class of morphological filters can be constructed that can often be used in place of the standard linear filters. Whereas linear

filters sometimes distort the underlying geometric form of an image, morphological filter leave much of that form intact.

In the following sections, definitions and properties of binary and gray-scale morphological transforms will be introduced. Examples of morphological operations with binary and gray-scale images will be shown throughout the sections.

Binary Morphology

The two most fundamental operators in mathematicalal morphology are dilation and erosion (Haralick and Shapiro, 1993). Almost all morphological operations can be defined in terms of these two basic operations. Dilation is the morphological transformation that combines two sets by using vector addition of set elements. Binary dilation was first used by Minkowski, and in the mathematicals literature it is called *Minkowski addition.* If B and K are sets in N-space (E^N) with elements b and k respectively, $b = (b_1,..., b_N)$ and $k=(k_1,...k_N)$ being N-tuples of element coordinates, then the dilation of B by K is the set of all possible vector sums of pairs of elements, one coming from B and one coming from K.

The *dilation* of B by K is denoted by $B \oplus K$ and is defined by

$$B \oplus K = \{c \in E^N | c = b + k \text{ for some } b \in B \text{ and } k \in K\} \qquad (2.1)$$

The first set B of the dilation $B \oplus K$ is associated with the image's underlying morphological processing, and the second set K is referred to as the *structuring element,* the shape that acts on B through the dilation operation to produce the result $B \oplus K$.

Stated more intuitively, the mathematicalal definition of dilation says that a foreground pixel will be written to the output set at all positions of the structuring element where any foreground pixel in the structuring element overlays a foreground pixel of the image set. The result of this operation is illustrated in Fig. 2.2. Notice how dilation fills in holes in the object and expands its boundaries, filling in any narrow 'creeks' that might exist.

(a) Original Image   (b) Dilation of (a)
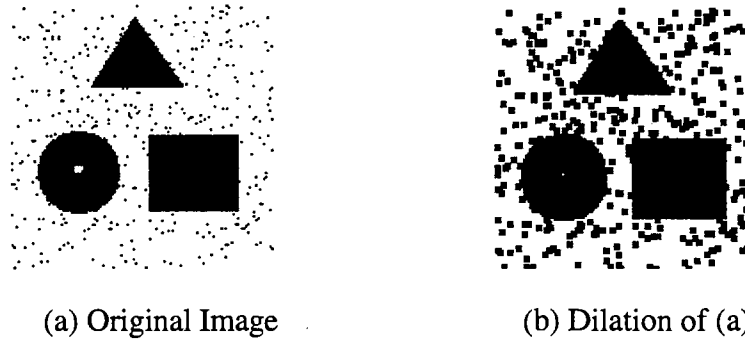
Fig. 2.2 Binary Dilation

Erosion is the morphological dual of dilation, that is, the dilation of a set B is equivalent of the erosion of the complement set B^*. It is the morphological transformation that combines two sets by using containment as its basis set. If B and K are sets in Euclidean N-space, then the erosion of B by K is the set of all elements x for which $x + k \in B$ for $k \in K$.

The erosion of B by K is denote by $B \Theta K$ and is defined by

$$B \Theta K = \{x \in E^N \mid x + k \in B \text{ for every } k \in K \} \qquad (2.2)$$

Erosion can be viewed as a morphological transformation that combines two sets by using vector subtraction of set elements. Expressed as a difference of elements b and k, erosion becomes

$$B \ominus K = \{x \in E^N | \text{ for every } k \in K, \text{ there existed a } b \in B \text{ such that } x = b\text{-}k\} \qquad (2.3)$$

An example of this operation is shown in Fig. 2.3. The eroded image of Fig. 2.3(b) is produced by stepping the structuring element, or template, over the input until all of the foreground pixels of the template fit over foreground pixels in the underlying image. At each position where this is true a pixel is written to the output array corresponding to the reference pixel position.
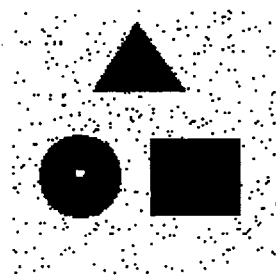
(a) Original Image   (b) Erosion of (a)

Fig. 2.3 Binary Erosion

Erosion enlarges holes in the object, shrinks its boundary, eliminates 'islands' and removes narrow 'peninsulas' that might exist on the boundary. The dilation and erosion transformations bear a marked similarity in that what one does to the image foreground the other does to the image background.

Besides dilation and erosion, there are two transformations, opening and closing, which Matheron (1975) also treated as fundamental transformations in that they can be constructed directly from dilation and erosion. Although dilation and erosion are dual operations, it is not possible to reconstruct an image set by the application of dilation after previously having eroded the image. The dilation operation will only be able to

reconstitute the essential features of the structure of the object as modified by the structuring element.

The *opening* of image B by structuring element K is denoted by $B \circ K$ and defined by

$$B \circ K = (B \ominus K) \oplus K \tag{2.4}$$

The *Closing* of image B by structuring element K is denoted by

$$B \bullet K = (B \oplus K) \ominus K \tag{2.5}$$

The results of these operations are illustrated in Fig. 2.4. Opening and closing operations form the basis of boundary smoothing and noise elimination processes whether the noise is manifested as small holes within the object or as small protrusions outside the object. Continuing with the geographical metaphor which seems so appropriate for the effect of morphological operations, opening smoothes object 'coastlines', eliminates small 'islands', and cuts narrow 'isthmuses'. Thus, it isolates objects which may be slightly touching one another.

Like erosion and dilation, opening and closing are dual transformations; an opening of the background of an object behaves like a closing of the object. The *open-close* filter is defined as an opening followed by a closing, *i.e. open-close* of image B by structuring element K is denoted by

$$B_k = (B \circ K) \bullet K \tag{2.6}$$

(a) Original Image   (b) Opening of (a)   (c) Closing of (a)
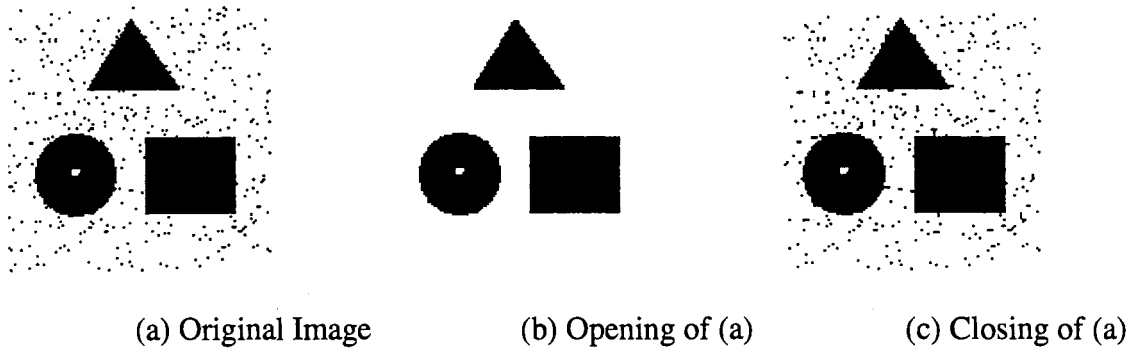
Fig. 2.4 Binary Opening and Closing

(a) Original Image   (b) Open-close of (a)   (c) Close-open of (a)
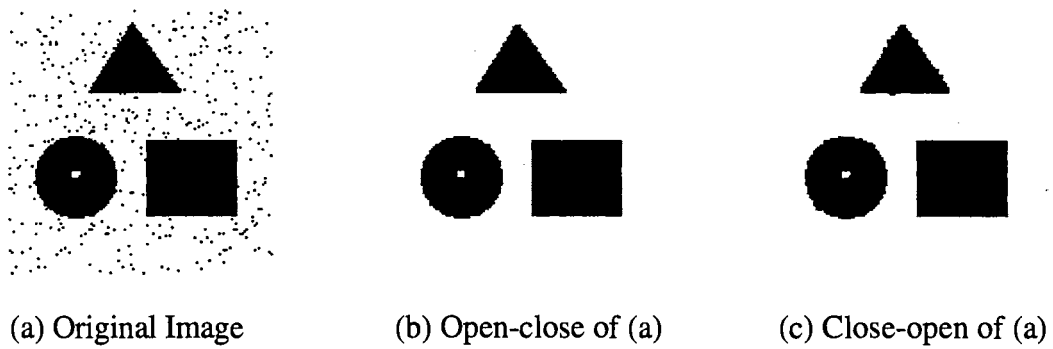
Fig. 2.5 Binary Open-close and Close-open

The reverse is true for the *close-open* operation, *i.e. close-open* of image B by structuring element K is denoted by

$$B^k = (B \bullet K) \circ K \tag{2.7}$$

Transformations that apply products of opening and closing are less severe and introduce less distortion than an individual closing or opening (Morales and Ko, 1996). Examples of *open-close* and *close-open* are shown in Fig. 2.5.

Sternberg (1986) introduced a new class of morphological filters called the alternating sequential filter (ASF), which is defined as an iterative application of openings and closings with different size structuring elements. The ASF is denoted by

$$ASF(B) = B_{Kn} B_{Kn-1} \ldots B_1 \tag{2.8}$$

where n is an integer, K_n, $K_{n-1}, \ldots K_1$ are structuring elements with different sizes, and $K_n \supseteq K_{n-1} \supseteq \ldots \supseteq K_1$. Schonfeld *et al* (1991) proved that ASF is the best in preserving crucial structures of binary images in the least mean difference sense.

The *open-close* filter is used in this research to generate the morphological pyramid for ATR. Applying open-close filters with identical structuring elements to the different size (resolution) of the image to generate a pyramid is equivalent to applying ASF to the image (Sternberg, 1986). A morphological pyramid can be denoted by

$$MP(B) = B_K^n B_K^{n-1} \ldots B_K^1 = ASF(B). \tag{2.9}$$

where K is the SE, $B^n, B^{n-1}, \ldots B^1$ are the original images with different resolution (size) through subsampling, and $B^n \supseteq B^{n-1} \supseteq \ldots \supseteq B^1$.

Gray Scale Morphology

The binary morphological operations of dilation, erosion, opening, and closing are all naturally extended to gray scale imagery by the use of a *min* or *max* operation.

A function, $f(x)$, dilated by a structuring element, K, is defined by

$$(f \oplus K)(x) = \max_{y \in K}\{f(x - y)\} \qquad (2.10)$$

where $x \in D$ in Z^2, Z is a set of integers, and K is a subset of Z^2. That is, the dilation is a moving local maximum operator. Similarly, the erosion, which is the moving local minimum operator, is defined by

$$(f \ominus K)(x) = \min_{y \in K}(f(x + y)\} \qquad (2.11)$$

The expressions for grayscale dilation and erosion bear a marked similarity to the convolution frequently encountered in image processing, with sums and differences replacing multiplication and minimum and maximum replacing summation.

For grayscale images, the erosion and dilation operations have the dual effect of eliminating positive-going impulse noise and negative-going impulse noise, respectively. However, each operation also biases the signal either upward or downward as shown in Fig.2.6. The dilated image (Fig. 2.6(b)) has highlighted, expanded bright regions, while the eroded image (Fig. 2.6(c)) has been biased negatively, producing a dark image, with expended dark regions.

(a) Original Image

(b) Dilation of (a)

(c) Erosion of (a)

Fig. 2.6 Grayscale dilation and erosion

Mathematicalally, the opening and closing are defined respectively as

$$f \circ K = (f \ominus K) \oplus K \qquad (2.12)$$

$$f \bullet K = (f \oplus K) \ominus K \qquad (2.13)$$

Apply an opening operation to an image will smooth the image, reject positive-going impulses, and preserve edge information as shown in Fig. 2.7(a). When applied to an image, the close operation will again smooth noise without removing edges, but will eradicate negative-going impulses as shown in Fig. 2.7(b). Fig. 2.7(a) and Fig. 2.7(b) also indicate that grayscale openings and closings by smooth structuring elements round off sharp grayscale distinctions.

The open-close and close-open filters are formed by the concatenation of open and close operations. The open-close filter is an opening followed by a closing while close-open is the reverse. Both of these operations have the ability to smooth noise, especially impulse noise, of both the positive-going and negative-going type, with little bias. So, I propose to use an open-close filter to construct the MP for ATR. The results of the open-close and close-open operations are shown in Fig. 2.8(a) and Fig. 2.8(b).

Properties of Morphological Filters

Morphological filters, including dilation, erosion, opening, and closing, are both increasing and translation invariant. The increasing property means that containment relationships are maintained through the morphological operation. For example, $A \subseteq B$

(a) Opening of Fig. 2.6(a)   (b) Closing of Fig. 2.6(a)

Fig. 2.7 Grayscale opening and closing

(a) Open-close of Fig. 2.6(a)   (b) Close-open of Fig. 2.6(a)

Fig. 2.8 Grayscale Open-close and close-open

implies $A \oplus K \subseteq B \oplus K$. If the translation of image A by x is denoted by $(A)_x$, then the translation invariance property means that $(A)_x \oplus B = (A \oplus B)_x$.

Besides being increasing and translation invariant, opening and closing are idempotent. By idempotency, we mean that the result of an opening or closing is unchanged when we reapply the opening or closing, $i.e.$ $(A°K)°K = A°K$. This corresponds to the notion of bandpass filtering in the linear filtering theory. An ideal bandpass filter is idempotent. Once we have removed all those frequencies in a signal outside of a given band, no further spectral changes are introduced by sending the signal through a second bandpass filter identical to the first.

These morphological filters are the scale generating mechanisms used in morphological pyramids. Now, the image pyramid will be discussed.

Multiresolution Pyramids

Multiresolution representations are important tools in computer vision for image filtering and analysis at multiple scales. The pyramid has been developed in the computer vision field as a general framework for implementing highly efficient algorithms, including algorithms for motion analysis and object recognition. The pyramid schemes have been traditionally utilized for region-based image segmentation (Burt and Rosenfeld, 1981), multiresolution edge detection (Acton, 1995; Chen and Acton, 1997), motion detection (Burt, 1989), signal reconstruction (Adelson et $al.$, 1987), and image coding (Burt and Adelson, 1983). They are also applied extensively in image compression and progressive

transmission (Burt, 1989; Burt and Adelson, 1983, Florencio and Schafer, 1994; Toet, 1989; Sun and Maragos, 1989; Cantoni and Levialdi, 1994), and promise excellent performance and flexibility (Burt, 1989; Cantoni and Levialdi, 1994).

Pyramidal architectures generate the same image at different resolutions and scales. The goal is to ensure the use of the most appropriate resolution for the operation, the task, and the image at hand. Furthermore, a hierarchical structure allows us to undertake the solution of problems at low spatial resolution and to proceed to successive refinements until the final verification of the results is reached at the highest resolution available. In this way acceleration factors greater than the number of processing elements (PEs) can be obtained even in the low-level vision tasks, because data reduction is an exponential function of the number of levels that have been constructed (Cantoni and Levialdi, 1994).

In a pyramid data structure, images are represented by a sequence of copies of the original data in which both sample density and resolution are decreased in regular steps, from the base (containing the original image) toward the apex (Fig. 2.9). The degree of reduction between successive image copies can vary over a wide range. In the most common solution, the number of pixels is reduced by at factor of 4 with each ascending level.

The pyramid construction from the original image can be accomplished in many ways. The simplest way consists of direct subsampling. Each level is obtained by keeping every other row and column from the previous level. Other, more sophisticated solutions are usually pursued in order to preserve or enhance some image properties.

Apex

Pyramid Level #3   Filtering, Subsampling

Pyramid Level #2   Filtering, Subsampling

Pyramid Level #1   Filtering, Subsampling

Scale Increases, Resolution Decreases

Pyramid Level #0

Original Image

Base
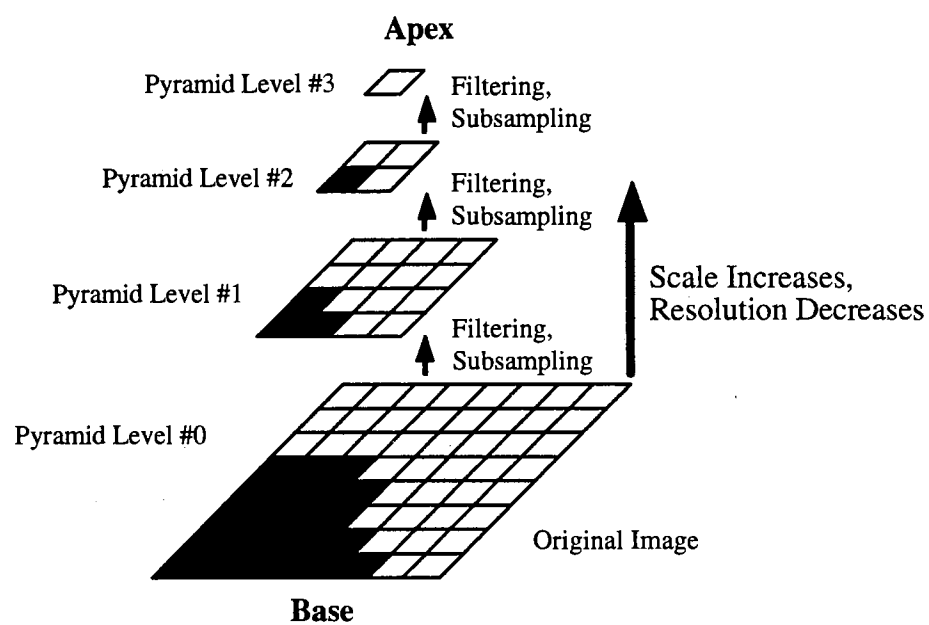
Fig. 2.9 The Structure of a Multiresolution Pyramid

The most common approach to pyramid construction is to apply linear filters with progressively increasing spatial extent to generate a sequence of images with progressively decreasing resolution. The low-pass filter not only creates a more coarse scale representation but also limits aliasing (distortion) from the subsampling operation. In the following section, two linear pyramid representations, Gaussian and Laplacian, pyramids will be analyzed.

Gaussian and Laplacian Pyramids

The Gaussian pyramid (GP) (Burt, 1988), which contains subsampled images that have been iteratively smoothed using a linear Gaussian-weighted filter, is by far the most prevalent pyramid in the literature. A GP can be constructed by convolving the image with a spatially localized generating function and then subsampling by a factor of 2 in each linear dimension. Linear generating functions that have been used for the GP include the Gaussian-weighted filter (Burt and Adelson, 1983) and the uniformly-weighted (average) filter (Wells, 1986).

The Gaussian pyramid represents images in multiresolution by a sequence of low-pass filtered images $\{G_0, G_1, G_2,..., G_n\}$. G_0 is the original image, which is convolved with a low pass filter w then subsampled by discarding every other row and column to form G_1. G_1 is then filtered and subsampled to form G_2, and so on. In general, for $l > 0$,

$$G_l = [w*G_{l-1}]_{\downarrow 2}.$$ (2.14)

Here the notation $[.]_{\downarrow 2}$ indicates that the image contained within the brackets is subsampled by a factor of 2 in each spatial dimension. The filter w is called the *generating kernel*. In practice this is chosen to be spacially limited and separable. Separable means the convolution can be decomposed into two successive one-dimentional convolutions: one operating on the image row by row, and the other operating on the image column by column. In this way, the computational cost of the filter convolution is kept to a minimum.

Fig. 2.10 depicts a 256x256 IR image of an airplane that is used as the initial image for GP. In Fig. 2.11, levels 1-3 of the GP are shown with the image size expanded to the size of the original image. The shortcomings that limit the effectiveness of linear pyramidal approaches to recognition and tracking are illustrated in this example. By level 2 of GP (64x64), the image of the aircraft has been severely blurred (Fig. 2.11(b)). At level 3 (32x32), the internal IR information for the target has been lost (Fig. 2.11(c)). Furthermore, the boundaries of the aircraft have been effaced.

The Laplacian pyramid is commonly used as the basis for multiresolution analysis in image coding (Adelson, *et al.* 1987). Roughly speaking, levels of the Laplacian pyramid are formed as difference images between successive levels of the Gaussian pyramid. If G_l is the lth level of a Gaussian pyramid, then the corresponding Laplacian level L_l is given by

$$L_l = G_l - w * G_l \qquad\qquad (2.15)$$

The Laplacian pyramid has a number of properties that make it appropriate for image coding. First of all, it is a complete image representation: the original image can be recovered exactly from its Laplacian pyramid representation through a simple inverse transformation. Moreover, it is a compact representation, and one that tends to enhance

Fig. 2.10 Original 256x256 IR image

(a) GP Level 1 (128x128)   (b) GP Level 2 (64x64)   (c) GP Level 3 (32x32)

Fig. 2.11 Three Levels of a Gaussian Pyramid

important features of an image, such as edges. Features are further separated over levels of the pyramid according to their scale. These properties also make the Laplacian pyramid well suited for image compression and computer graphics (Adelson, *et al.* 1987). However, the poor edge localization of the Laplacian pyramid limits its application to ATR.

Linear pyramidal techniques have also been utilized for object recognition and tracking (Burt, 1989; Cantoni and Levialdi, 1994). With linear pyramids, the coarse-to-fine tracking strategy was first explored. For a successful coarse-to-fine guided search, image features of potential targets must be first identified on a coarse pyramid level by target region boundaries (edges) and the incorporated intensity information. Using the GP, edges can be located on each level by evaluating the local image gradient or by locating zero-crossings in the Laplacian of the Gaussian (LoG) pyramid images. Unfortunately, the linear filtering leads to an increased disparity between image edges and the actual physical boundaries between pyramid levels for a coarse-to-fine search. The effects of region bleeding/merging will also degrade recognition performance and reliability. With the linear generating function, the target edges are oversmoothed, not enhanced.

More recently, nonlinear pyramidal techniques for image segmentation and edge detection have appeared. With a nonlinear method, the linear generating function is replaced by nonlinear operation. Such a substitution may render the pyramid unsuitable for such tasks as image subband coding or signal reconstruction, but may prove useful for feature extraction and segmentation in terms of feature preservation in the coarse pyramid levels. Acton (1994) used a diffusion-based generating function to preserve edge

localization and internal radiometric properties of the image regions. Sun and Maragos (1989) applied morphological pyramids to image compression and progressive transmission. Toet (1989) used MP for image decomposition. The resulting image description provided a useful basis for multiresolution shape analysis. Moreover, the MP image decomposition scheme is well suited for VLSI implementation. Also, Florencio and Schafer (1994) used a shape-preserving morphological filter to generate coarse image representations in a pyramid for image coding.

One might ask the question "There are other efficacious nonlinear filters, such as the well-known median filter. Why choose a morphological filter to construct the pyramid?" The morphological filter can preserve the critical features of the image, and is more computationally efficient than other nonlinear filters. In the following, the relationship between median filter and morphological filter is discussed in order to answer this question.

Morphological Filters Compared to Median Filters

It is shown in (Maragos and Schafer, 1987) that median filtering of any signal by convex windows is bounded below by openings and above by closings. In addition, the open-close and close-open filters suppress impulse noise similar to the median, can discriminate between positive and negative noise impulses, and are computationally less complex than the median.

The relationship between median filter and morphological open-close is stated in the following theorem.

Theorem (Maragos, 1987) The open-close and close-open filters with a window B of any finite extent function f are roots of the median by a window W. Here the sizes of window B and W are $|B|=n+1$ and $|W|=2n+1$, and n are positive integers.

A root of a filter is any signal such that $\Psi(f) = f$ so that the result of the filtering is unchanged and stable. The above theorem can be formulated as

$$f \bullet B \circ B = \underbrace{med[med[...med[f,W]...]]}_{m} \qquad (2.16)$$

$$f \circ B \bullet B = \underbrace{med[med[...med[f,W]...]]}_{m} \qquad (2.17)$$

where m is the smallest number that the result of the iteration of the median filter is stable. The above equations indicate that the open-close and close-open filters yield a median root in one pass while median filter needs iterate several times to yield a median root. Therefore, the morphological open-close and close-open filters are more efficient in eliminating impulse noise than the median filter.

To understand this theorem intuitively, examples are given. Fig. 2.12(a) shows the 256x256 Cameraman image corrupted by salt-and-pepper noise with variance = 5. In Fig.2.12(b), (a) is opened by a 2x2 square convex window B. It shows that opening can suppresses the positive noise ("salt" noise). In Fig.2.12(c), the open-close suppresses the negative noise ("pepper" noise) remained in Fig. 2.12(b). Fig.2.12(d) is the median filtering result of Fig.2.12(a) by a 3x3 convex square window W. Comparing Fig.2.12(c) and Fig.2.12(d) indicates that a median filtering of an image by a 3x3 window behaves

(a)Salt-and-pepper noise (σ^2=5)

(c) open-close (2x2)

(b)open (2x2)

(d) Median filter (3x3)

Fig. 2.12 Comparison of Morphological Filters to Median Filters

similarly to the open-close by a 2x2 window *B*, but the latter is computationally less expensive than the median.

This section has discussed two linear multiresolution pyramids: the Gaussian Pyramid and the Laplacian Pyramid. Their limitations for ATR are indicated. The relationship between two nonlinear filters, the median filter and the morphological filter, is also discussed, and demonstrates the advantage of using nonlinear morphological filters to construct pyramids. The applications of the nonlinear pyramids are also briefly presented. At this time, the MP has been applied to image coding, segmentation, image representation, and image restoration. However, no report has been found that solves the ATR problem with a morphological pyramid. This thesis addresses the utilization of the morphological pyramid structure to automatically identify targets.

CHAPTER III

MORPHOLOGICAL PYRAMIDS FOR ATR

Chapter Overview

So far, existing ATR techniques and the background of the morphological pyramid have been introduced. Although the morphological pyramid has been applied to image compression, image representation, and image restoration, there has been no application to automatic target recognition. In this chapter, the generation of optimal morphological pyramids for ATR will be discussed.

The morphological image pyramid is a collection of images derived from a given image with decreasing sizes and resolutions as the Gaussian pyramid. Images at each level are derived from previous level by filtering and subsampling. For morphological pyramids, filtering is typically accomplished by a morphological *opening*, *closing*, or *open-close*, which can be treated as a low-pass filter. The subsampling is usually carried out by discarding every other row and column from the previous level.

Mathematically, given an image I and a structuring element K, a morphological pyramid can be defined as a collection of images, MP = $\{I_L, L= 0, 1,..., N\}$, where I_L is the subsampled image at level L from the previous filtered image, and N is the largest integer

for which I_L is non-empty. The procedure for computing this pyramid can be summarized as follows:

(1) Initialization: $I_0 = I$

(2) $I_L = [(I_{(L-1)} \Diamond K]\downarrow s$

(3) repeat (2) until $L = \log_s(n)$

Here $\Diamond$ denotes the morphological operation and n is the image size. $[]\downarrow s$ indicates the image contained within brackets is subsampled by a factor of S in each spatial dimension.

From the previous equations, it is clear that the appropriate morphological filter, the size of SE in the morphological filter, and the sampling rate need to be chosen to construct an MP which preserves the critical information between pyramidal levels. In the following sections, these three choices are discussed.

Selection of the Structuring Element

Since the performance of morphological operations is highly dependent on the size and shape of the structuring element, the SE used to construct MP needs to be investigated. If the size of the SE is too large, the morphological filter will remove some important features of targets whose size is smaller than the SE. Thus, the probability of target detection will decrease. On the other hand, if the size of the SE is too small, the noise in the background whose size is larger than the SE can not be effectively reduced.

One characteristic of mathematical morphology is that it requires an exact specification of the structuring element. Many applications of mathematical morphology

use only a single structuring element. In some cases, this may not produce the best result. For example, a structuring element with a disk shape or a square shape is used in most morphological filtering algorithms which results in the elimination of some geometrical details in the image along with the noise. One way to avoid the blurring effect is to use a smaller structuring element. However, using smaller structuring elements will not remove noise effectively. Sternberg (1986) introduced the iterative algorithm using multiple structuring elements. In his morphological filter, openings and closings are applied iteratively using multiple structuring elements which are spheres of different diameters. It has been shown that the use of a complex morphological filter with multiple structuring elements can remove noise and preserve geometrical features more effectively than a simple morphological filter. By applying same size structuring element in the different levels in a pyramid, we actually apply the different size SE to the original image.

Fig. 3.2 shows that for a binary image corrupted with impulse noise, the use of a complex binary morphological filter with only two structuring elements can remove noise effectively and preserve more geometrical features than a simple one SE morphological filter. Since the two structuring elements shown in Fig. 3.1 match the directions of the camera control line, this control line is preserved in Fig. 3.2(c) and (d). These results also demonstrate that with the same SE, applying open-close is better than simply applying open.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Fig.3.1 The two 3x3 SE's used in the filter of Fig. 3.2

(a) Noisy Cameraman

(b) Opening of (a) using
one 3x3 square SE

(c) Opening of (a) using
the two SE's in Fig. 3.9
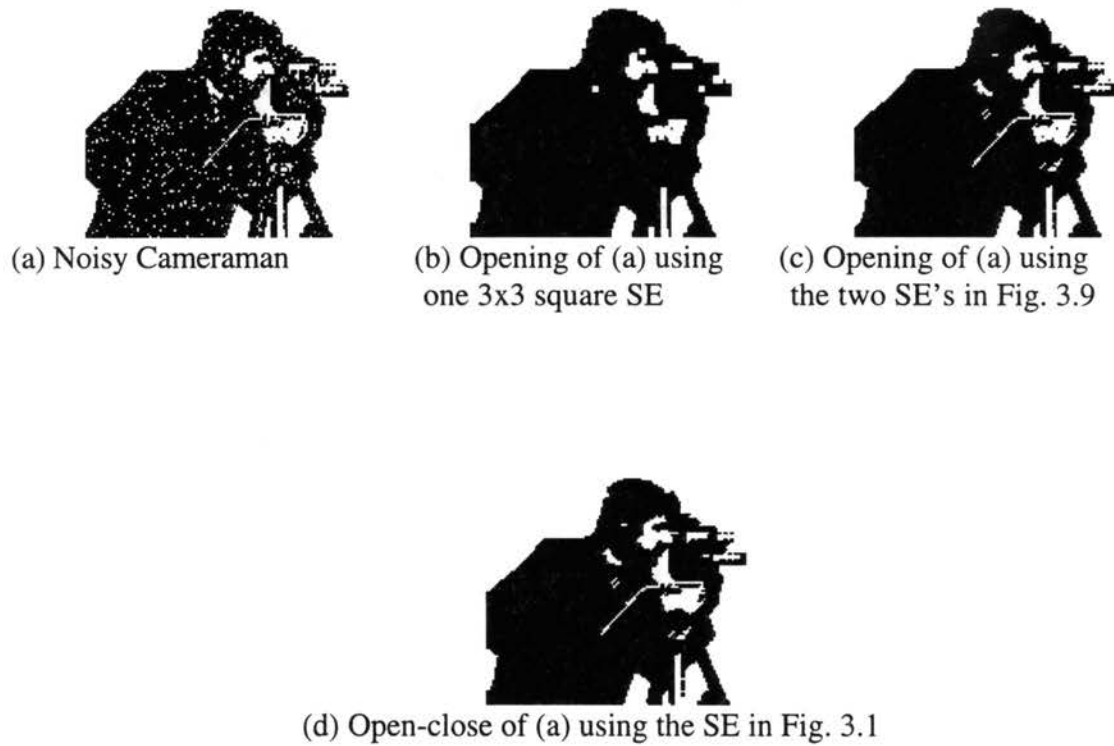

(d) Open-close of (a) using the SE in Fig. 3.1

Fig.3.2 Illustration of selection of the SE

The choice of SE involves the selection of two features: shape and size. The shape of the SE is discussed in the following. The size of the SE is described in a later section.

The optimal shape of the SE should match the directions of the target features so that the shape of the target is maximally preserved as shown in Fig. 3.1-3.2. Since the optimal shape of the SE depends on the shape of the object of interest in the scene, the shape of the object needs to be analyzed and defined. Before the mathematical definition of the target is given, two definitions--- *convex* set and *piecewise convex* set, need to be given.

Convex Set and Target Shapes

A convex set is a collection of points having the property that for any two points in the set, a line segment joining them is also completely within the set. Fig. 3.3(a) shows a convex set. Fig. 3.3(b) shows a non-convex set.

(a)   (b)
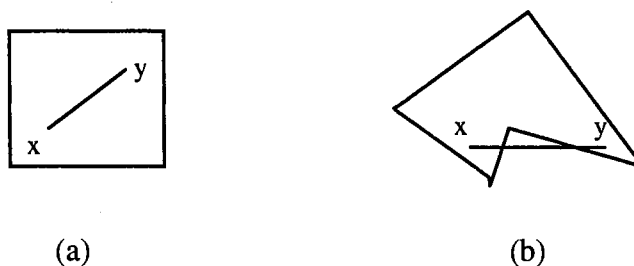
Fig. 3.3 Convex Set

A piecewise convex set is a union of convex sets that share elements only at boundaries. For example, the set in Fig. 3.4 is piecewise convex set. In Fig. 3.4 both A_1

and A_2 are convex sets, and they are connected and share the boundary e. Mathematically, the piecewise convex set A can be defined as

$$A = A_1 \bigcup_e A_2 .$$ (3.1)

$$e = A_1 \cap A_2 .$$ (3.2)

where $\bigcup_e$ denotes a disjoint union with shared boundary e.
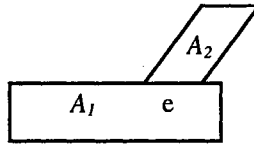
Fig. 3.4 Piecewise Convexity

Now the shape of the target can be expressed by the following equation:

$$A = \bigcup_{e_i} A_i \; : e_i \neq \varnothing, \text{ and } 1 \leq i \leq n$$ (3.3)

where A_i is a convex set, and n is a bounded integer. The target can then be described as a piecewise convex set with joint boundaries.

Selection of the SE Shape

To preserve the features of the target, structuring elements should be chosen such that SE's match the direction of the target features. The procedure of selecting the SE shape is as follows:

(1) Let $E_0^i = A_i$. (3.4)

$$(2)\ E_1^i = E_0^i \ominus B \tag{3.5}$$

$$E_j^i = E_{j-1}^i \ominus B \tag{3.6}$$

$$(3)\ N = \min(j): E_j^i = \varnothing \tag{3.7}$$

$$(4)\ S_i = E_{N-2}^i \tag{3.8}$$

where N is an integer, B is a symmetric structuring element (please note that this structuring element is different from the structuring element K which is used to construct MP), E_j^i is the ith part of a target under jth iteration of erosions and $E_0^i \supseteq E_1^i \supseteq \ldots E_N$. S_i is the minimum shape of the ith part of a target A_i and is the ith SE shape. The example shown in Fig. 3.5 demonstrates this procedure. In this example, $S_i = E_{N-2}^i = E_1^i = E_0^i \ominus B$ which is Fig. 3.5(c). This is the optimal shape of SE because it preserves the shapes of the object of interest.

Then, the multiple structuring element is a disjoint union of S_i

$$SE = \bigcup_{e_i} S_i \ . \tag{3.9}$$

The selection of the optimal SE shapes has been discussed above. Now, the optimal SE size needs to be chosen. The optimal size of the SE should be selected such that the connection between parts of the object of interest is preserved. To choose the optimal SE size in construction the MP, the homotopy preserving critical sampling theorem (HPCST) is utilized. Before introducing the HPCST, we need to give a definition of homotopy.

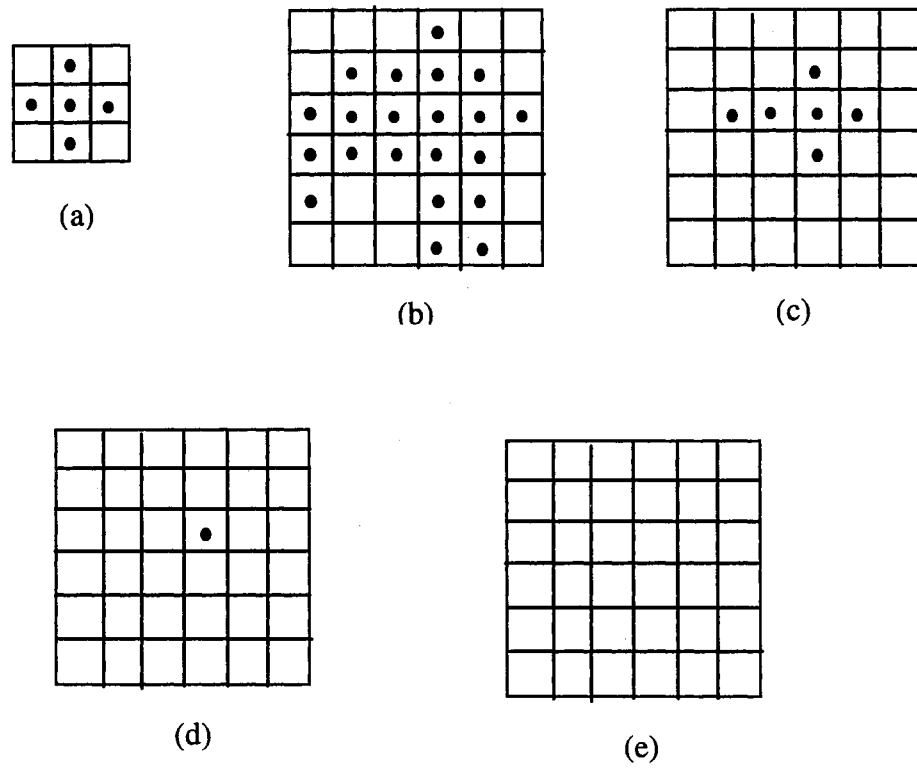Fig. 3.5 Generation of the minimum shapes of a target A_i .

(a)SE B; (b) $E_0^i = A_i$; (c) $E_1^i = E_0^i \ominus B$; (d) $E_2^i = E_1^i \ominus B$; (e) $E_3^i = \varnothing$.

Since $N = \min(j): E_j^i = \varnothing$, N=3 in this case.

Sampling and Homotopy

A morphological pyramid is a hierarchy of images where copies of the original image are morphologically filtered and sampled in regular steps. The sampling process in generating the MP is different from the conventional sampling process. The conventional sampling process is a mapping from continuous signals to digital signals. On the other hand, the sampling process used to construct the MP is performed in the discrete time domain, and can be accomplished by discarding a number of rows and columns in the image. Therefore, it is refereed as subsampling. The goal in subsampling is to preserve the integrity of the target information, and eliminate the insignificant details in each pyramid level.

The integrity of objects can be quantified by *homotopy*. Two sets are homotopic if a one-to-one correspondence exists between objects in two image representations. For example, Fig. 3.6(a) and (b) are homotopic sets. Fig. 3.6(a) and (c) are not homotopic sets. The evaluation of the homotopy for the MP will prevent region merging and allow target/background discrimination for coarse representations.

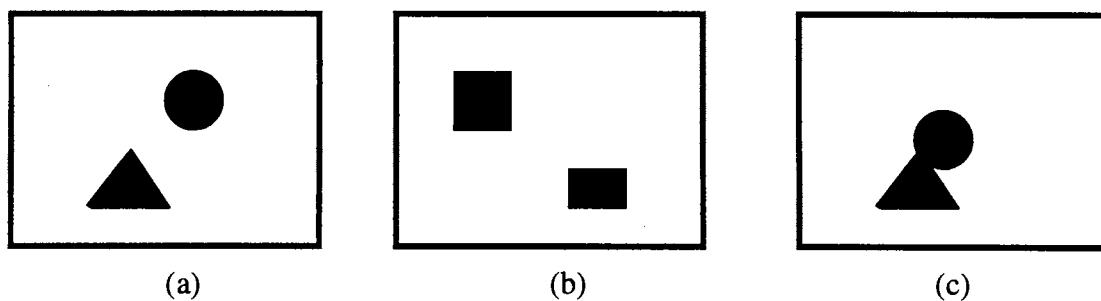(a)   (b)   (c)

Fig. 3.6 Homotopy

It is not straightforward to choose a subsampling scheme that guarantees the preservation of the homotopy. The traditional sampling theorem often serves as the basis for analysis of the *aliasing*. Aliasing occurs when the frequency components of the sampled signal are distorted by overlapping of the original spectra. This sampling theorem can be used for designing the down-sampling lowpass filter for anti-aliasing in linear systems. To generate an image pyramid, the traditional sampling theorem does not seem to be appropriate where contour and equivalent information, not the frequency content, are of primary concern. Furthermore, the morphological filter is a nonlinear filter whose frequency response can not be analyzed. Therefore, a sampling theorem based on morphology needs to be utilized.

It is well known in linear signal processing that prevention of aliasing can be accomplished by low-pass filtering the signal according to Shannon's sampling theorem. To construct a morphological image pyramid, the traditional sampling theorem is not pertinent, since the morphological filter is a nonlinear filter. The frequency response of this nonlinear filter cannot be analyzed. In this case, the morphological sampling theorem (MST) is utilized to design the anti-aliasing filter (Haralick *et al*, 1989).

The MST describes how to remove small objects, object protrusions, object intrusions, and holes before sampling to prevent a sampled result that is unrepresentative of the original one. In linear signal processing, the presence of frequencies higher than the Nyquist frequency causes the distortion in the sampled signal. Likewise in morphology, the presence of small details must be filtered to prevent distortion.

The subject of morphological sampling has been addressed by several authors (Florencio and Schafer, 1994; Heijmans, 1991; Haralick *et al*, 1989). Heijmans and Toet (1991), using a specific definition for sampling ("probing element sampling"), presented a very general sampling strategy, which includes results on general sampling grids as well as on regular grids. With a more traditional sampling definition, Haralick *et al.* (1989) provided many relations between the sampled signal and the original signal under several conditions.

The purpose of subsampling is to reduce the amount of data used to represent the original signal. The sampling process should be such that as much useful information as possible is retained from the original signal. A desirable property of the sampling process is that it should be invertible, *i.e.* it should be possible to *reconstruct* an original image from its sampled version. Of course, this cannot be done in general since we have lost information by discarding a number of rows and columns in the subsampling process. However, it is possible to give conditions in order to make the sampling process invertible. The morphological sampling theorem (Haralick *et al*, 1989) describes how a digital image must be filtered in order to preserve relevant information. It indicates to what precision a morphologically filtered image can be reconstructed after sampling. It specifies the relationship between the original image and its morphologically filtered and sampled one. It permits us to construct an optimal morphological pyramid for ATR. An optimal morphological pyramid should satisfy the following conditions: (1)It should maximally remove noise and unwanted details, (2)It should preserve the shapes of the objects of interest, and (3)It should preserve homotopy between two adjacent pyramid levels for the

object of interest. In the following, the morphological sampling theorem will be briefly discussed.

Morphological Sampling Theorem

There are four basic sets in the morphological sampling theorem: the original set F, the sampling set S, the structuring element K, and the reconstruction structuring element C. Before presenting the morphological sampling theorem, the definitions of sampling and reconstruction must be given.
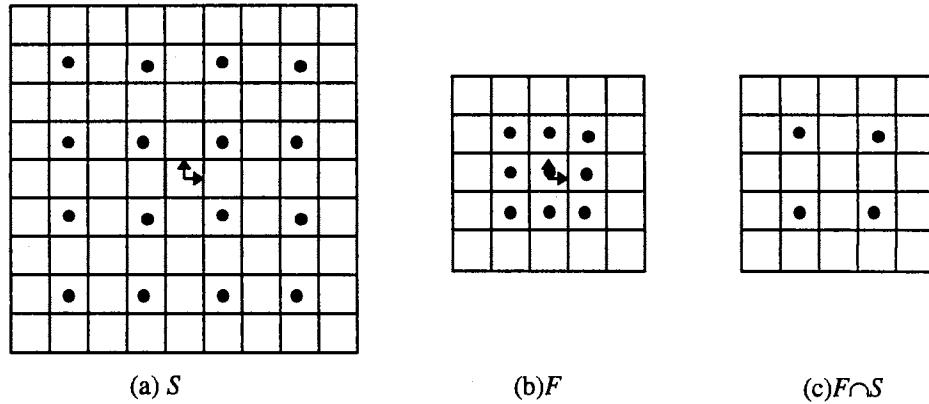

(a) S (b)F (c)$F \cap S$

Fig. 3.7 Illustration of sampling process

The sampling operator $\sigma(.)$ is defined as:

$$x \in \sigma(F) \Leftrightarrow x \in F \cap S. \tag{3.10}$$

Thus, by (3.10) sampling a set is equivalent to taking the intersection of the original set F with the sampling set S. Fig. 3.7 illustrates the sampling process. The

sample set S is chosen as the set of even numbers in both row and column. The original set F is shown in Fig. 3.7(b). Fig. 3.7(c) is the sampling result. Here ↳ denotes the origin.

This example shows that if the sample points of S are too finely spaced, little will be accomplished by the reduction in resolution. For instance, the sampling set is selected as the set of every pixel by row and by column. The result of sampling set $F \cap S$ is the same as the original set F. On the other hand, if S is too coarse, objects may be missed by the sampling. For example, the sampling set is chosen as the set of every fifth pixel by row and by column. In this case the result set $F \cap S$ is empty.

There are two kinds of reconstructions of the sampled images: a maximal reconstruction accomplished by dilation and a minimal reconstruction accomplished by closing. The reconstruction operator $\rho(.)$ is defined as:

$$\rho(G) = G \oplus C \quad \text{or} \quad \rho(G) = G \bullet C \tag{3.11}$$

In morphological sampling theorem, the sampling set S and the reconstruction SE C must satisfy following conditions:

$$S \oplus S = S \tag{3.12}$$

$$S = \breve{S}, \tag{3.13}$$

$$C \cap S = \{0\}, \tag{3.14}$$

$$C = \breve{C}, \tag{3.15}$$

and
$$a \in C_b \rightarrow C_a \cap C_b \cap S \neq \varnothing \tag{3.16}$$

where $\breve{}$ denotes reflection with respect to the origin, and C_a is the translation of C by a. Conditions (3.12) and (3.13) mean that S is an infinite set, and it is symmetric. Conditions

(3.14) and (3.15) imply that C is symmetric and is small enough so that its intersection with S is $\{0\}$. Condition (3.16) states that the sampling space must be small enough such that the radius of C contains two sample intervals but not three. Fig. 3.8(b) illustrates a structuring element C satisfying the conditions (3.14-3.15).
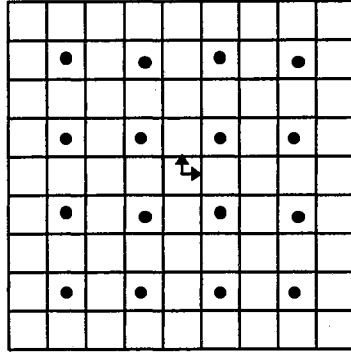
(a)Sampling every other pixel by row and column. The sampling set S is represented by all points which are shown as • .
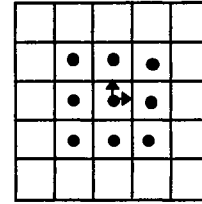
(b)C

A symmetric element C. For the sampling set S of Fig. (a)

$C \cap S = \{0\}$

Fig. 3.8 Reconstruction SE Conditions

Before the four propositions of the morphological sampling theorem is presented, the definition of aliasing needs to be given. Aliasing means that the reconstructed set is a distorted version of the original set. It can be expressed as:

$$\rho(\sigma(F)) \neq F .$$

(3.17)

The morphological sampling theorem consists of four propositions. The first proposition in this theorem shows that if we sample the reconstructed set, we obtain the same result as if we sample the original set. This is very important, since it shows that

after the first reconstruction, the signal belongs to a class of "perfectly reconstructible sets" associated with the sampling and reconstruction transformations.

Proposition 1 (Haralick *et al*, 1989)

$$\sigma(F) = \sigma(\rho(\sigma(F)))$$
(3.18)

This proposition is a key point in designing an appropriate anti-aliasing filter. Since this class of signals can be perfectly reconstructed, it is obvious that these signals should be preserved by the filtering.

The next proposition shows that the reconstructed set is close to the original set-- it differs from the original by at most a dilation or erosion.

Proposition 2 (Haralick *et al*, 1989)

$$F \ominus \check{C} \subseteq \rho(\sigma(F)) \subseteq F \oplus C$$
(3.19)

Looking from the other side, the next proposition shows that signals that were appropriately filtered can be estimated by the sampled signal.

Proposition 3 (Haralick *et al*, 1989)

$$\text{If } F = F \bullet C = F \circ \check{C} \text{ then } \rho(\sigma(F)) \ominus C \subseteq F \subseteq \rho(\sigma(F)) \oplus \check{C}$$
(3.20)

Notice the contrast between Propositions 2 and 3. While the first uses the original signal to put bounds on the sampled signal, the latter uses the sampled signal to put bounds on the original signal.

Proposition 4 [Haralick *et al*, 1989]

If $C \subseteq K,$ then

$$[(F \circ K \bullet K) \cap S] \bullet C \subseteq F \circ K \bullet K \subseteq [(F \circ K \bullet K) \cap S] \oplus C.$$
(3.21)

$$d[F \circ K \bullet K, (\{F \circ K \bullet K\} \cap S) \bullet C] \leq r(C).$$
(3.22)

$$d[(\{F \circ K \bullet K\} \cap S) \oplus C, F \circ K \bullet K,] \le r(C). \tag{3.23}$$

where $r(C)$ denotes the radius of C, $d[E,F]$ denotes the distance between E and F.

This theorem can be seen as the morphological sampling theorem equivalent of the Shannon Theorem. According to the Shannon Theorem, frequencies higher than the Nyquist rate cannot be reconstructed. Similarly, the morphological sampling theorem indicates that the position accuracy smaller than the radius of the reconstruction structuring element cannot be reconstructed. While the Shannon Theorem establishes a class of signals (closed under linear combinations) that can be sampled and perfectly reconstructed, the morphological sampling theorem establishes a class of signals that can be sampled and perfectly reconstructed. The morphological pyramids satisfy this theorem are the optimal ones for target identification.

In the following, the morphological sampling theorem is applied to choose the optimal morphological filter for the MP.

Choice of the Optimal Morphological Filter for the MP

The optimal morphological filter is a morphological filter which removes the noise and unwanted details but maximally preserves the object shapes and critical features. To generate the optimal MP for ATR, the desired property is that the sampling process can be inverted. In other words, the information in the sampled image is maximally preserved compared to the unsampled image. Proposition 3 in MST states that the sampled image is close to the original image if $F = F \bullet C = F \circ \check{C}$. Normally this condition can not be met

for an unprocessed image. Recall that the open and close operations are idempotent. If an image is opened by C once, reapplying the opening operation would not change the result. In this case, the opened image is equal to original one: $F = F \circ \breve{C}$. Similarly, if an image is closed by C once, reapplying the closing operation would not change the result, i.e. $F = F \bullet C$. Therefore, an image opened by C and closed by C (open-closed by C) meets the condition $F = F \bullet C = F \circ \breve{C}$. This image is the image which belongs to the class of perfectly reconstructible images.

Proposition 4 in MST gives the relationship of the open-close filter in the sampled domain and unsampled domain. It also states that the difference between the reconstructed and open-close image is not greater than the radius of reconstruct SE C. Since C is a subset of SE K, this difference is not greater than the radius of SE K. Therefore, the difference between the reconstructed and open-close image is determined by SE K. Now, applying this proposition to our problem, when the MP satisfies the condition of $F = F \bullet K = F \circ K$, it also meets the condition $F = F \bullet C = F \circ \breve{C}$ in proposition 3. This theorem supports the choice of the open-close filter for construction of the optimal MP for ATR.

The open-close filter is chosen as the anti-aliasing filter in this work. Small objects and object protrusions can be eliminated by a suitable opening, and small object intrusions and holes can be eliminated by a suitable closing (Burton and Benning, 1981). The open-close operation is the concatenation of the two filters. By carefully selecting structuring element sizes and shapes, openings and closings can be made which selectively remove image features according to their sizes and orientations. Furthermore, by applying

openings and closings iteratively, it makes no difference whether we are treating the foreground or the background of an image because openings and closings are dual transforms. Finally, iterative openings and closings of grayscale images filter out image noise without adding grayscale bias (Sternberg, 1986).

The open-close pyramid can be constructed by

$$I_1 = [I_0 \circ K_1 \bullet K_1]_{\downarrow 2} \qquad (3.24)$$

$$I_2 = [I_1 \circ K_2 \bullet K_2]_{\downarrow 2} \qquad (3.25)$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$I_N = [I_{N-1} \circ K_N \bullet K_N]_{\downarrow 2} \qquad (3.26)$$

The structuring elements K_i used with different levels in the pyramid can be either identical or different.

Homotopy Preserving Critical Sampling Theorem and SE Size

The optimal morphological filter and the shape of the structuring element of the morphological filter used to construct MP have been determined in the previous sections. In this section, the size of the SE which preserves the homotopy needs to be decided. The Homotopy Preserving Critical Sampling Theorem (HPCST) is utilized to determine the SE size (Florencio *et al*, 1994). The following is the HPCST:

Theorem Let $X \in P\{R^2\}$, let X^c be its complement, and let $(\sigma(X), D)$ and $(\sigma(X^c), D)$ be their respective planar graph representations induced by a rectangular lattice S with spacing a

and 6-way connectivity (Fig. 3.9). $K(r)$ is the SE with radius r. $\sigma(.)$ is the sampling operator, $\rho(.)$ the reconstruction operator, and K is the structuring element. Then the following hold:

1. If $X = X \bullet K(r) = X \circ \breve{K}(r)$ for some $r = a\sqrt{2}$, then the planar graph representations preserve the homotopy of X and X^c.

2. The condition in 1) above is not enough to guarantee the homotopy preservation if $r < a\sqrt{2}$.
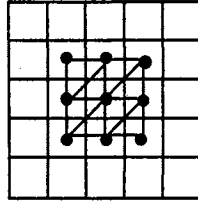
Fig. 3.9 6-way connectivity (The center point is 6-way connected to its neighbors)

HPCST states that if the sampling rate is S, the diameter of the structuring element must be greater than or equal to S multiply by $\sqrt{2}$ to preserve homotopy between the two successive levels. In this work, we use S equal to 2, therefore, the diameter of the SE will be 3. The SE used to construct MP can be as follows:

$$K = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The equations (3.23-3.24) are used to construct the MP. The first three levels of the MP are shown with the image size expanded to the size of the original image in Fig.

3.10, using Fig. 2.10 as the initial image. The morphological operation is open-close. The pyramidal images are generated with identical SE's.

It can be observed that the integrity of the target (airborne) and the connection between several parts (land, clouds, and target) of the image in MP have been preserved after repeated filtering and subsampling compared with GP (Fig. 2.11).
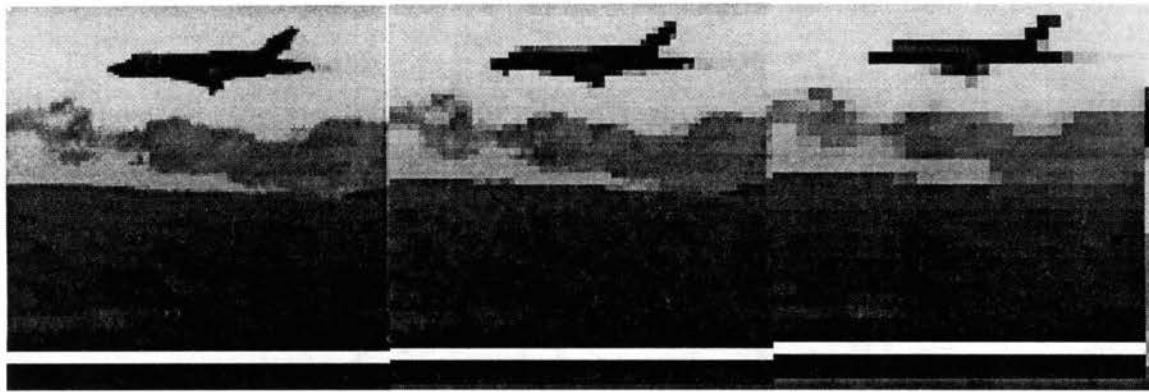
(a) MP Level 1 (128x128)   (b) MP Level 2 (64x64)   (c) MP Level 3 (32x32)

Fig. 3.10 Three Levels of a Morphological Pyramid

It is not enough to just preserve the homotopy between the two successive pyramid levels. The homotopy between the original image and the level for the initial searching---*root level,* needs to be preserved. This issue is discussed in the following section.

Selection of the Root Level

In our coarse-to-fine search algorithm, we must determine the root level L_r in MP which is between level 0 to level N for initial search. The root level is the level at which the targets have not vanished and the target's homotopy is maintained. The determination of the root level is a critical selection since the instance of a target missed at this level will never be recovered. In general, the selection obviously depends on the original spatial resolution and the maximum shift in position allowed to the target in the search space. Here, the morphological filtering at coarser levels can be quantified by region homotopy. The evaluation of the homotopy for the morphological pyramid (MP) will prevent region merging and allow target/background discrimination for coarse (sparsely sampled) representations.

If the root level is chosen to be too high, such as level 7 for a 256x256 image, the information of a target would disappear because there is only one pixel in the level 7 with the sampling rate S equal to 2. On the other side, if the root level is too low, such as level 0, the advantage of the pyramid structure for identification would not be utilized. In the following, we will give the analysis of root level choice.

The root level depends on the size of the target object and the minimum distance between two objects. Recall that an object is a union of convex sets with shared boundaries. Let a be the minor axis of the largest convex set of the object. In level L, the length of the minor axis will become:

$$a_{L} = \frac{a}{S^{L}}. \tag{3.27}$$

Where S denotes the sampling rate. The object will disappear when

$$\frac{a}{S^{L}} < S. \tag{3.28}$$

Therefore, the level in which the object before it disappears can be described as

$$L_{a} = \frac{\log_{10} a}{\log_{10} S} - 1. \tag{3.29}$$

Similarly, if the minimum distance between two objects is d, when the level reaches

$$L_{d} = \frac{\log_{10} d}{\log_{10} S} - 1. \tag{3.30}$$

The two objects would just before merge. Thus the root level is determined by

$$L_{r} = \min \{ L_{a}, L_{d} \}. \tag{3.31}$$

For example, if the minor axis a of the largest convex set of the object of interest equals 16, the sampling rate is selected to be 2 and the minimum distance d between the two objects is 8. According to Eq. (3.29-3.31), $L_{d} = 2$ and $L_{a} = 3$. Therefore the root level L_{r} will be level 2 in the pyramid. In this example, level 2 preserves the homotopy between original image and the initial searching level. Using SE size 3, sampling rate 2, and root level 2, not only is the homotopy between successive pyramid levels preserved, but also the homotopy between the original image and the level for initial searching is preserved.

In summary, a morphological pyramid has been defined that meets the criteria of optimality.

CHAPTER IV

OBJECT IDENTIFICATION AND TRACKING BASED ON MORPHOLOGICAL
IMAGE PYRAMIDS

Chapter Overview

The theoretical analysis of the generation of the optimal MP for target identification and tracking has been discussed in chapter III. In current chapter, the utilization of the nonlinear morphological pyramid structures for automatic target identification and tracking is presented.

Pyramidal image representations and algorithms are particularly important in the application of computer vision to object recognition. To achieve real-time processing a vision system must quickly locate potential objects of interest within a scene, then efficiently extract critical features that permit object recognition. Systems with fixed resolution spend more computational time on irrelevant image details, and require minutes or hours to identify an object that can be located in a second with the pyramidal structure (Burton and Benning, 1981).

The block diagram of this MP-based ATR algorithm is shown in Figure 4.1. The first step is to construct morphological image pyramids which will be described in the
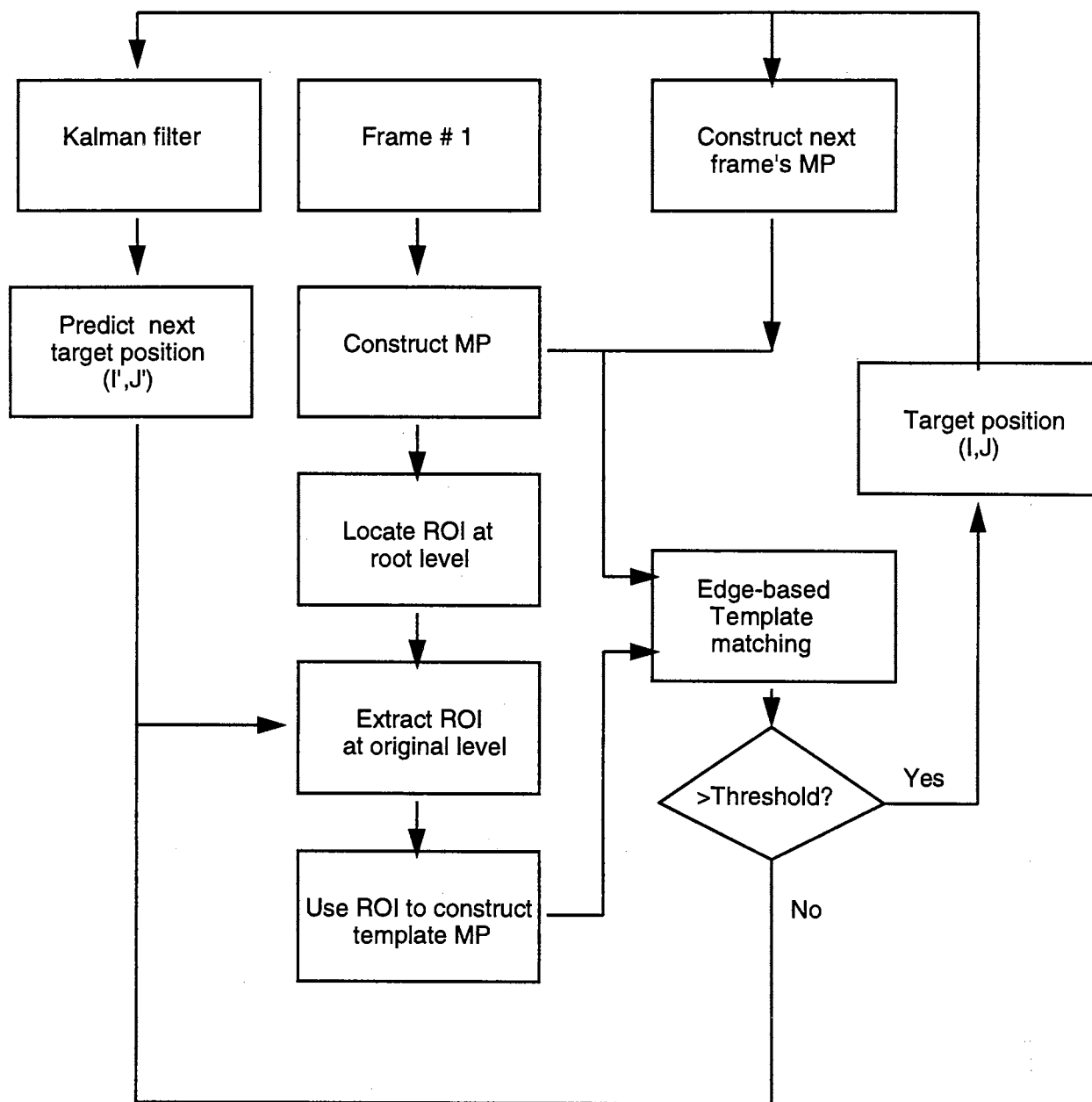
Fig. 4.1 ATR Flowchart

following section. The next step is to segment the image by utilizing the MP structure and node link technique. Step 3 is to locate and extract the region of interest (ROI) where a candidate target is present. Step 4 is to construct the template morphological pyramids using the ROI as the original image and the same technique used to generate image pyramids. Step 5 is to perform a coarse-to-fine search using the template matching technique between the image pyramids and template pyramids. Finally, the integrating the MP and Kalman predictor is addressed to further improve the performance of the system. The following sections will discuss the details of these steps.

Construction of the MP

Recall that the fundamental operators in morphology are dilation, erosion, opening and closing [Chapter II]. A grayscale function, $f(x)$, dilated by a structuring element, K, is defined by

$$(f \oplus K)(x) = \max_{y \in K}\{f(x-y)\}. \tag{4.1}$$

where $x \in D$ in Z^2, Z is a set of integers, and K is a subset of Z^2. Similarly, the erosion is defined by

$$(f \ominus K)(x) = \min_{y \in K}(f(x+y)\}. \tag{4.2}$$

The opening and closing are defined respectively as

$$f \circ K = (f \ominus K) \oplus K, \tag{4.3}$$

and $\quad f \bullet K = (f \oplus K) \ominus K. \tag{4.4}$

In construction of the MP for edge detection, morphological open-close, which is an opening followed by a closing, is used for filtering images. The advantage of using morphological open-close is that it is increasing, translation invariant, and idempotent. Idempotency guarantees that the result of an open-close is unchanged after re-application, corresponding to bandpass filters in linear signal processing. Moreover, open-close can maximally preserve the shape of targets and eliminate noises.

Given an image I and a structuring element K, a morphological pyramid can be defined as a collection of images, $MP = \{I_L, L = 0, 1,..., R\}$, where I_L is the subsampled image at level L computed from the previous filtered image at level L-1, and R is the root level. The procedure for constructing a pyramid with an open-close filter can be summarized as follows:

(1) Initialization: $I_0 = I$

(2) $I_L = [(I_{(L-1)} \circ K \bullet K]_{\downarrow S}$

(3) repeat (2) until $L=R$.

Here the notation $[.]_{\downarrow S}$ indicates the image contained within brackets is subsampled by a factor of S in each spatial dimension (along both rows and columns).

The MP-Based Segmentation

The technique used for identification of targets is the coarse-to-fine template matching which is discussed in later sections of this chapter. Template matching is an operation that can be used to find out how well a template sub-image matches a window

of a given image. The coarse-to-fine template matching can be performed on the original gray scale images or on binary images. The latter one, though, is much faster. Therefore, binary template matching is utilized in our coarse-to-fine search in order to speed up the searching process. One of the classic techniques to transform a gray scale image to a binary image is segmentation.

Segmentation is the process of isolating each object from the rest of the scene. There are two categories of segmentation: (1)edge detection and (2)region growing. The edge detection approach focuses on finding discontinuities in intensity variations or pattern variations. The edges form closed boundaries which define regions. Region growing based schemes focus on finding contiguous pixels which have similar intensities or patterns. In this way groupings of homogeneous parts of the image are formed.

Thresholding is a particularly useful region-approach segmentation technique for scenes containing solid objects on an even background. But, in many cases, the background gray level is not constant, and the contrast of objects varies within the image. In such a case, the thresholding technique works poorly.

The classic gradient-based edge detectors are sensitive to noise and produce more false edges than morphological edge detectors (Ramesh and Haralick, 1992). Recently, the multiresolution approaches have attracted attention because of their computational efficiency and their robustness in the presence of noise. The popular linear Gaussian pyramid is not well-suited for edge detection due to the smoothing effect of Gaussian convolution. A nonlinear pyramid, the anisotropic diffusion pyramid (ADP), edge detector was presented in (Acton, 1996) which showed decreased edge localization error,

compared to the linear Gaussian-based pyramid edge detector. Edge detection utilizing morphological pyramids is attractive because they can be efficiently implemented, in contrast to the ADP, and the MP offers the same low edge localization error as the ADP.

The images in this system are segmented by using a pyramid node linking technique (Burt, 1981) on the MP. A child-parent relationship is defined between nodes (pixels) in adjacent levels. For each node in level l there are 4x4 candidate children nodes at level l-1 and 4 candidate parent nodes at level l+1. Links between adjacent levels in the pyramid are formed based on similarities. Each node is linked to a single one of these candidate parents. After linking, the pixel values are re-computed as follows:

$$I_l(i,j) = I_{l+1}(\text{int}(\frac{i}{2}) + a, \text{int}(\frac{j}{2}) + b).$$
(4.5)

$$(a,b) = \underset{(m,n)}{\arg\inf}\left|I_{l+1}(\text{int}(\frac{i}{2}) + m, \text{int}(\frac{j}{2}) + n) - I_l(i,j)\right|$$
(4.6)

where $(m,n) \in \{(0,0),(0,1),(1,0),(1,1)\}$. This process repeats until the pixel values of level 0 remain unchanged.

The algorithm utilizes a multiresolution pyramidal structure created by successive morphological operations and subsampling of the original image. The boundaries detected at a low resolution level of the morphological pyramid (MP) are used to guide the detection of discontinuities at higher resolution levels. Through linking, segments are formed in level 0 from the coarse-to-fine segmentation originating at level R. From the segmented image, a binary image is obtained by thresholding, and an edge map is generated by locating the boundaries between the segmented regions.

Fig. 4.2(a) shows an original 256x256 image of a desk surface with a pair of scissors and two pens corrupted by additive Laplacian-distributed impulse noise with variance 27.8. The morphological operation used to construct MP is close, the structuring element used here is a 3x3 square window, and the sampling rate is $S=2$.

The performance of the MP-based edge detector is compared with another multiresolution edge detector---the anisotropic diffusion pyramid (ADP) based edge detector reported in (Acton, 1996). The ADP gives the edge map shown in Fig. 4.2(b) using Fig. 4.2(a) as the original image. The edge map created by utilizing MP structure is shown in Fig. 4.2(c) using the same original image. Both methods yield excellent edge localization, and create continuous, thin edge contours that reflect the structurally significant objects in the scene. Note that the MP created connected regions for both the scissors and the pens while the ADP generated gaps in these objects.

The computational complexities of the MP and the ADP are in stark contrast. For one level of the MP, the number of comparisons (adds) for the closing is $2M^2P^2$, where M is the size of structuring element (=3 typically), and $P \times P$ is the image size. For one level of the ADP, there are $8kP^2$ floating point (f.p.) adds, $8kP^2$ f.p. multiplications and $4kP^2$ f.p. exponential operations involved, where k is the number of diffusion iterations. Here a typical of $k \approx 10$ (Acton, 1994). Clearly, the construction of the ADP is more expensive than creation of the MP.

Fig. 4.3 compares the MP-based edge detector with the traditional gradient-based method. Fig. 4.3(b) is the edge map given by gradient edge detector using Fig. 4.3(a) as the original image. Fig. 4.3(c) is the edge map detected by the MP-based method. It

shows the MP-based method yields superior edge localization, and creates clean and continuous edge contours while gradient method generates noise and broken edges.

The segmentation and resulting edge detection yielded by the MP is particularly effective in the presence of noise. The experimental results demonstrate superior solution quality over standard full resolution detectors and over previous pyramidal approaches. Because of the low computational cost of the MP edge detector, it is suitable for video tracking, image and video compression, and real-time target recognition.
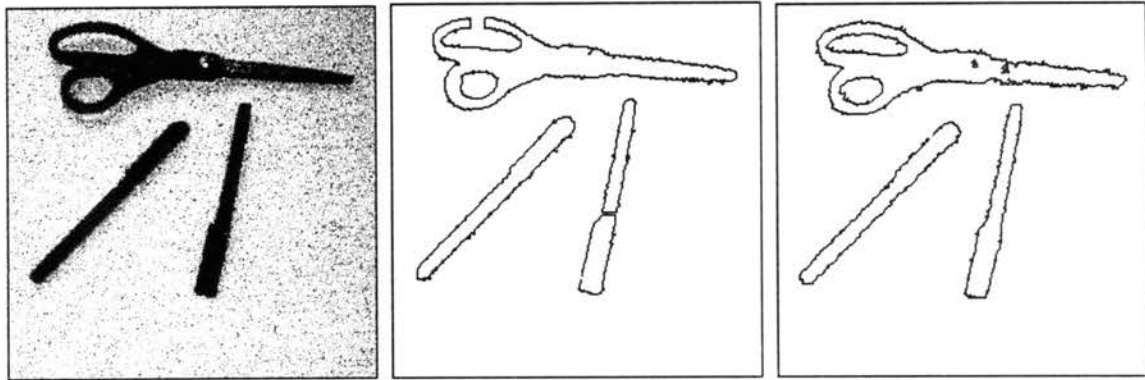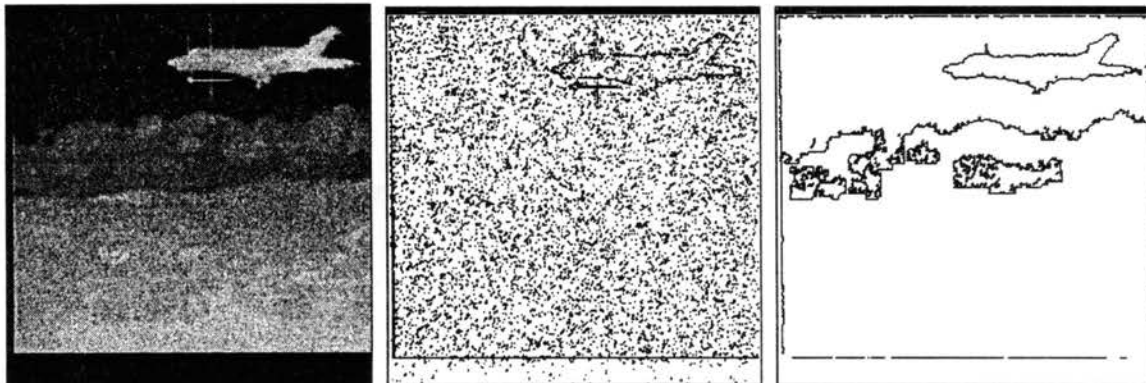
(a)Original image   (b)Edges from the ADP   (c) Edges from the MP

Fig. 4.2 Edges from the ADP and the MP

(a)Original image   (b)Edges from Gradient   (c) Edges from the MP

Fig. 4.3 Edges from Gradient and the MP

Initial Detection of the Target

In this section, the location of the region of interest (ROI) is presented. The ROI is the region where a candidate target is present. By utilizing the MP structure, the target area can be extracted easily. The adaptive thresholding technique is applied to segment the potential target at a coarser level which will provide the rough location of the candidate target. Once the result of this first screening of the ROI is obtained, the second phase of refinement has to be performed. The refinement is accomplished step by step at each intermediate level until the maximum resolution level (the original image) is reached. At each level the confidence of the selection is checked, and the spatial position of the target is refined by checking the g nearest neighbors corresponding to the position of the previous level.

After the objects are segmented, the connected component labeling is used to label all the objects in the scene. The size of all the objects are calculated in this program. Assuming the background is the largest object in the scene and the interesting object is the second largest object in the scene, the identification of the interesting object is straightforward: the object whose size is the second largest one will be the target.

Coarse-to-fine Template Matching and Target Detection

In the following, the object identification algorithm based on morphological pyramid structure which utilized adaptive template matching scheme (Pratt, 1991) is described.

Template matching is one of the most fundamental means of object detection in a scene. The template is a replica of an object of interest within the scene. In the template matching process, the template is sequentially scanned over the scene and the common region between the template and the scene is compared for similarity.

A template match is rarely ever exact because of image noise. The template matching can be conducted by measuring the difference between the template and the image at all points of the image. The usual difference measure is the mean-square error defined by

$$D(m,n) = \sum_j \sum_k [I(j,k) - T(j-m,k-n)]^2 \tag{4.7}$$

Where $I(j,k)$ denotes the image field to be searched and $T(j,k)$ is the template. Expansion of (4.7) yields

$$D(m,n) = D_1(m,n) - 2D_2(m,n) + D_3(m,n) \tag{4.8}$$

Where

$$D_1(m,n) = \sum_j \sum_k [I(j,k)]^2 \tag{4.9}$$

$$D_2(m,n) = \sum_j \sum_k I(j,k)T(j-m,k-n) \tag{4.10}$$

$$D_3(m,n) = \sum_j \sum_k [T(j-m,k-n)]^2 \qquad (4.11)$$

The term $D_3(m,n)$ represents a summation of the template energy and is a constant value independent of the coordinate (m,n). The term $D_1(m,n)$ represents the image energy over the window area and generally varies slowly over the image field. The term $D_2(m,n)$ is the cross correlation between the image field and the template. This term should become large to yield a small difference at the coordinate of a template match. However, the cross correlation is not always an adequate measure of the template match because the image energy term $D_1(m,n)$ is position variant. For example, the cross correlation can become large, even under a condition of template mismatch, if the image amplitude over the window area is high at a particular coordinate (m,n). This difficulty can be avoided by the normalized cross correlation

$$C(m,n) = \frac{D2(m,n)}{D1(m,n)} = \frac{\sum_j \sum_k I(j,k)T(j-m,k-n)}{\sum_j \sum_k [I(j,k)]^2} \qquad (4.12)$$

The normalized cross correlation has a maximum value of unity that occurs if and only if the image function under the template exactly matches the template.

One of the major limitations of template matching is that it is very costly in computation. For a window of size M x M and an image size N x N, the computational complexity is $O(N^2M^2)$. Although fast correlation techniques such as fast Fourier transforms decrease the correlation computation, these techniques still require computations at each of $(N-M+1)^2$ locations. In our MP based coarse-to-fine template matching approach, a logarithmic efficiency can be achieved compared to the traditional

methods. Another way to speed up the searching process is to use binary template matching instead of the gray scale one.

Two MPs are created in our coarse-to-fine searching scheme, one for the template and the other for the search scene. Potential targets are initially located by searching at lowest resolution at considerably lower cost. In this case the search is formulated as the correlation of a low resolution template with a low resolution level of the image pyramid. The cross correlation at level L is formulated as:

$$C_L(m,n) = \frac{\sum_j \sum_k I_L(j,k)T_L(j-m,k-n)}{\sum_j \sum_k [I_L(j,k)]^2} \tag{4.13}$$

where $I_L(j,k)$ denotes the image field at level L of the MP and $T_L(j,k)$ is the template at level L of the template pyramid.

The searching process starts at the root level R. Assuming there is only one target in the scene, the position which has the maximum C_R is the target location. Once the target is found in the coarsest level R, the search procedure moves to the next higher resolution pyramid level and examine details that should occur within the target. Only the selected locations are searched at the next higher level. Therefore, the number of search positions is reduced to $q\log(N-M+1)^2$ where q is a constant.

Computation time is reduced in this procedure by organizing the search from coarse-to-fine. Burt has shown that when used in search applications, these multi-scale structures can have a significant computational advantage (Burt, 1989). In fact, a speedup of $2^{2(n-1)}$ is possible even if the search for a target is performed everywhere in each level,

where n is the number of scales (levels) in the pyramid structure. For example, if 3 scales are used for searching the target, the speed up is 16.

Integrating the MP and Kalman Predictor

There are several reasons to integrate a Kalman filter with the MP detector. The main goal is to provide further computational efficiency and increase the detection rate. Considering the following situations: the change of size, rotation, and occlusion of the target. The normalized cross correlation would fall low and matching is not possible in those cases. To enhance our multiresolution search system, integration of the MP and Kalman predictor is necessary.

The system integrating of the MP tracker and Kalman filter can be constructed easily. The system is initialized by a multiresolution search which locates the object of interest and passes its initial location to the Kalman predictor. The MP tracker continue to track the target. At the same time, each instance of the target located by this MP tracker is fed to a Kalman filter which predicts the location of target in the next frame. Once the matching between template and image falls below a certain threshold, because of target scale, target rotation, or change of viewpoint, a new template can be generated using the ROI guided by Kalman predictor.

In our pyramidal-based target identification, potential targets are initially located by searching at lowest resolution. In this case the search is formulated as the correlation of a low resolution template, or matched filter, with a low resolution level of the image

pyramid. Once a candidate target is found, the search procedure moves to the next higher resolution pyramid level and examines details that should occur within the target. The instance of the target of frame n located by this pyramidal matched filter is fed to a Kalman filter which predicts the location of the target in the frame $n+1$. This pyramidal tracker continues to track the target of frame $n+1$. The instance of the target in frame $n+1$ located by MP tracker are compared with Kalman predictor results. If the difference between those two locations is larger than a threshold, the result of MP tracker needs to be re-examined. In this case, the Kalman predictor acts as a quality check, and makes the detection more accurate.

In our application, the target or the camera may be partially occluded by dirt and vegetation. In such situations, the MP tracker is not able to find a matching no matter how you adjust the orientation and the scale of the template. The target is lost by the MP tracker. But, if the MP tracker and Kalman filter are allowed to interact, the position of the target can be located by Kalman predictor.

Kalman Filter (Kalman,1969; Sage and Melsa, 1982)

The Kalman filter is the optimal linear filter in the least mean square error sense for estimation of a future state based on the past and current states. It can be utilized to predict the target location based on its previous location. The Kalman filter is the combination of a predictor and a filter. The predictor estimates the location of the target at time t given $t-1$ observations and the filter improves this estimation by accounting for

measurement uncertainty and random drift. The Kalman filter is divided into two independent and identical systems for image tracking -- one corresponds to the horizontal i and another corresponds to the vertical j direction. Since these systems are identical, the following discussion will only present the Kalman filter for the i direction.

Given a sufficiently rapid acquisition rate, a constant velocity model named the α-β model of the Kalman filter can be used, and the changes in velocity can be modeled as a random drift. Then, the prediction of the next position of the target is given by

$$\hat{i}_{t+1|t} = \hat{i}_{t|t} + \delta T \, \hat{v}^i_{t+1|t} \tag{4.14}$$

where $\hat{i}_{t|t}$ is the filtered estimate of the observed location i_t at time t and $\hat{i}_{t+1|t}$ is the prediction for time $t+1$ at time t, $\hat{v}^i_{t+1|t}$ is the predicted velocity, and δT is the time difference in seconds between two observations (typically, 1/30 second). The filtered estimate $\hat{i}_{t|t}$ provides a filtered representation of the observed location, accounting for measurement uncertainty, and is given by

$$\hat{i}_{t|t} = \hat{i}_{t|t-1} + \alpha_t (i^0_t - \hat{i}_{t|t-1}), \tag{4.15}$$

where i^0_t is the observed location, $\hat{i}_{t|t-1}$ is the predicted location at time t - 1, and α_t is a filter gain. Note that the observation i^0_t contains measurement noise: $i^0_t = i_t + n_t$, so a filtered version $\hat{i}_{t|t}$ is utilized in track prediction. The gain α_t determines the balance between the previous track history and the new observation. If α_t is large (near 1), then the observations are very reliable, and the track history is ignored. In the case of strong measurement noise, α_t is set to a small value (near 0).

The target velocity v_t^i is modeled by

$$v_{t+1}^i = v_t^i + u_t^i \tag{4.16}$$

where u_t^i is a velocity drift process (acceleration). Then, the prediction for the velocity

component at time t is

$$\hat{v}_{t+1|t}^i = \hat{v}_{t|t-1}^i + \beta_t (i_t^0 - \hat{i}_{t|t-1}) \tag{4.17}$$

where β_t is also a filter gain. β_t controls the effect of the new observation on the

predicted velocity. If β_t is near 0, the observations are unreliable and the actual velocity

is fairly constant.

Computing the Kalman Gains

Assuming stationary, the Kalman filter gains (α_t and β_t) can be computed before

the tracker is implemented. Since these gains converge quickly to constants, only a few

computations are necessary. Both gains depend upon the variances of the noise processes

and the state vector error covariance matrix.

Let the state vector X_t be defined as $X_t = \begin{bmatrix} i_t \\ v_t^i \end{bmatrix}$, a column vector containing the

actual position and velocity of the target. The state vector of the predictor is then

expressed as $\hat{X}_{t|t-1} = \begin{bmatrix} \hat{i}_{t|t-1} \\ \hat{v}_{t|t-1}^i \end{bmatrix}$, and the state vector for the filter $\hat{X}_{t|t}$ is constructed

similarly as $\hat{X}_{t|t} = \begin{bmatrix} \hat{i}_{t|t} \\ \hat{v}_{t|t}^i \end{bmatrix}$. The error in the predicted state vector is $X_t - \hat{X}_{t|t-1}$, and the

error for the filter state vector is $X_t - \hat{X}_{t|t}$. Since these errors are stochastic vectors, they have covariance matrices. The predicted state vector covariance matrix is

$$P_{t|t-1} = E\left[\left(X_t - \hat{X}_{t|t-1}\right)\left(X_t - \hat{X}_{t|t-1}\right)^T\right],$$ (4.18)

and the filtered state vector error covariance matrix is

$$P_{t|t} = E\left[\left(X_t - \hat{X}_{t|t}\right)\left(X_t - \hat{X}_{t|t}\right)^T\right].$$ (4.19)

In order to provide the minimum mean squared error prediction, the α_t and β_t are chosen to minimize $P_{t|t}$.

Assuming normality for both noise processes, with measurement noise variance σ_n^2 and velocity drift noise variance σ_u^2, the Kalman solution that minimizes $P_{t|t}$ is

$$\alpha_t = \frac{P_{t|t-1}^{11}}{P_{t|t-1}^{11} + \sigma_n^2}$$ (4.20)

and

$$\beta_t = \frac{P_{t|t-1}^{21}\delta T}{P_{t|t-1}^{11} + \sigma_n^2}.$$ (4.21)

For the constant velocity α-β model, $P_{t|t-1}$ in (4.24) and (4.25) can be computed recursively as follows:

$$P_{t+1|t}^{11} = P_{t|t-1}^{11} + 2P_{t|t-1}^{12} + P_{t|t-1}^{22} - \frac{\left(P_{t|t-1}^{11} + P_{t|t-1}^{12}\right)^2}{P_{t|t-1}^{11} + \sigma_n^2},$$ (4.22)

$$P_{t+1|t}^{12} = P_{t|t-1}^{12} + P_{t|t-1}^{22} - P_{t|t-1}^{12}\left(\frac{P_{t|t-1}^{11} + P_{t|t-1}^{12}}{P_{t|t-1}^{11} + \sigma_n^2}\right),$$ (4.23)

$$P_{t+1|t}^{21} = P_{t+1|t}^{12},$$ (4.24)

$$P_{t+1|t}^{22} = P_{t|t-1}^{22} + \sigma_u^2 - \frac{\left(P_{t|t-1}^{12}\right)^2}{P_{t|t-1}^{11} + \sigma_n^2} \ . \tag{4.25}$$

Initializing the Kalman Filter

To implement the Kalman filter, the initial conditions for $P_{t|t-1}$ ($= P_{1|0}$) are needed.

For a description of the tracker initial conditions, two additional parameters: σ_i^2, the

variance in the initial position and $\sigma_{v_i}^2$, the variance of the initial velocity are defined.

Assuming that the initial position is a uniformly distributed random variable over the N

($N=256$ in our case) possible positions, the computation of σ_i^2 is straightforward. The

calculation of $\sigma_{v_i}^2$ can be derived in the same way using $a\ priori$ knowledge of the target's

minimum and maximum velocities, assuming that the initial velocity is also uniformly

distributed.

The filtered state vector error covariance at time 0 is given by

$$P_{0|0} = E\left[\left(X_t\right)\left(X_t\right)^T\right] = \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_{v_i}^2 \end{bmatrix} . \tag{4.26}$$

$P_{1|0}$ can be computed using

$$P_{1|0} = A_0 P_{0|0} A_0^T + Q_0 \tag{4.27}$$

where A_0 is the state transition matrix, and

$$A_0 = \begin{bmatrix} 1 & \delta T \\ 0 & 1 \end{bmatrix} . \tag{4.28}$$

Q_0 is the covariance of the system noise, and is given by

$$Q_0 = \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix}. \tag{4.29}$$

To initiate the tracker, $\hat{i}_{0|0}$ is set to the first acquired position. The first velocity estimate is indeterminate and can be set to any constant in the range of possible velocities.

Implementing the Kalman Filter

The Kalman filter is implemented using the five following steps:

(1) Compute the initial $P_{1|0}$ using (4.27).

(2) Use (4.20) and (4.21) to obtain and store the gains (α_t and β_t) for each t.

(3) Acquire the target position within the entire image using the MP.

(4) Use (4.15) to compute the filtered position $\hat{i}_{t|t}$, then (4.17) to obtain the predicted velocity $\hat{v}_{t+1|t}^i$. Finally, (4.14) is utilized to acquire the predicted position $\hat{i}_{t+1|t}$. (Repeat for the corresponding j terms).

(5) Acquire target within a track gate centered at predicted position using the MP. (If the target is lost at frame t, the track is "coasted" and i_t^0 is set to $\hat{i}_{t|t-1}$.) Return to step (4).

In this section, the MP-based identification and tracking system is integrated with a Kalman predictor. The Kalman filter has two functions in this system. One function is to check the detection quality. The other function is to provide the ROI for the new template when the occlusion happens.

CHAPTER V

COMPARATIVE RESULTS

This chapter illustrates applications of the morphological pyramid to ATR. Traditional ATR systems incorporate time-consuming pre-processing, detection, segmentation, and classification at a fixed spatial resolution on the original image. These processing algorithms are not feasible for a real-time implementation. This chapter presents the simulation results from using the MP-based identification and tracking technique to quickly and accurately locate a target, such as a fighter jet.

For each sequence tested, the proposed morphological pyramid-based identification and tracking method is applied, and the traditional fixed resolution method and Gaussian Pyramid method are also applied to the same image sequences for comparative purposes. In addition to testing the identification and tracking algorithms on an uncorrupted image sequence, the methods are also tested on noisy sequences with various signal-to-noise ratios to test the robustness of the proposed algorithm. Image noise from acquisition and transmission and clutter such as dust, smoke, and man-made clutter are inevitable. Therefore, it is necessary to test the algorithm under the conditions of noise.

In the following, the parameters for generation of the morphological pyramid used in ATR are given first. Then, the comparison results between the Gaussian pyramids based identification and tracking scheme and the MP-based method are presented. The computational time among the fixed resolution method, the GP technique, and the MP approach are also compared. Next, the parameters for Kalman predictor are given, followed by a comparison between the predicted results and ground truth data. Finally, some conclusions about the results are made.

Generation of Morphological Pyramids for ATR

Selection of Parameters for Morphological Pyramids

In chapter III, the theoretical analysis provides that the optimal morphological filter used in constructing the morphological pyramid is the open-close filter. Therefore, the morphological open-close operators are used to generate the MP for identification in conjunction with a one to two sampling scheme ($S=2$).

The size of the SEs is selected according to the homotopy preserving critical sampling theorem (HPCST): The diameter r of SE must be greater than or equal to the sampling rate S multiplied by $\sqrt{2}$ to preserve the homotopy between two successive levels. Therefore $r = S\sqrt{2} \approx 3$. For the example target provided in this chapter, the shape of the structuring element is determined by (3.3-3.9) which gives the SE as follows:

$$SE = S_1 \bigcup_e S_2 .$$

$$e = S_1 \bigcap S_2 .$$

$$S_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad S_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

In this case, structuring element S_1 preserves the shape of the fuselage, and S_2 preserves the shape of the tail (see Fig. 5.2).

Fig. 5.1 shows the pyramid constructed for the first frame of the sequence using the open-close filter, the defined SE, and the sampling rate $S=2$. Fig. 5.1(a) is the actual morphological image pyramid. Fig. 5.1(b) shows the same morphological pyramid with the image size expanded to the size of original image to provide a detail view.

An Example Sequence

Fig. 5.2 shows one representative example of a sequence of 25 infrared images of a fighter jet in flight. The original infrared images have a resolution of 256x256 pixels with 256 intensity levels in each pixel.

To display robustness of the MP, the simulations are performed on the same set of images but corrupted by Gaussian distributed noise. Figure 5.3(a) shows the morphological pyramid constructed for the first frame of the noisy sequence. The signal to noise ratio (SNR) of the test images is approximately 7dB. For comparison, the Gaussian pyramid for the same original image is shown in Fig. 5.3(b) using a 3x3 square
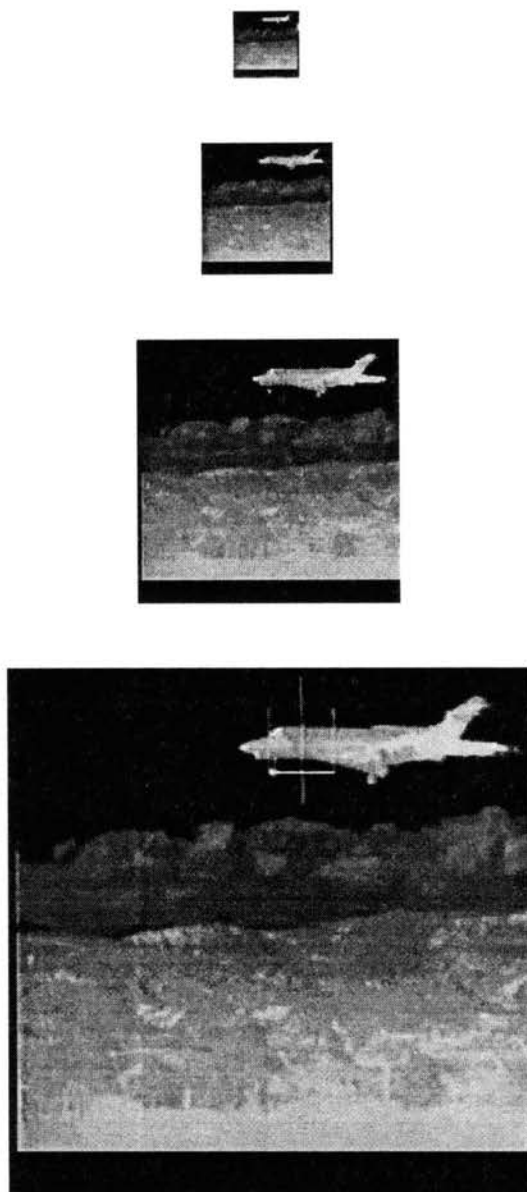
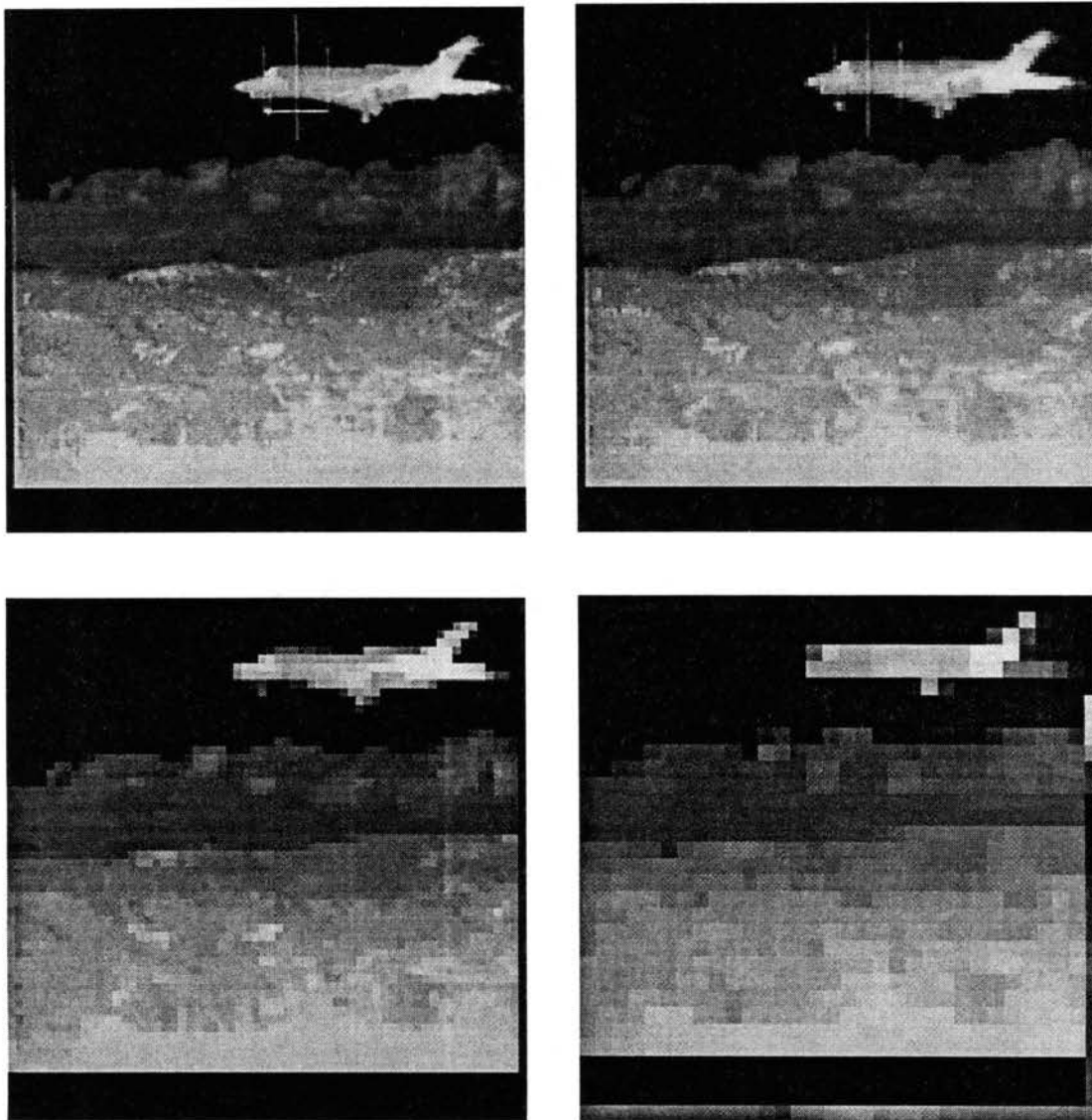Fig. 5.1(a) The MP for frame #1 of the original sequence

Fig. 5.1(b) The MP for frame #1 of the original sequence with the image size expanded to

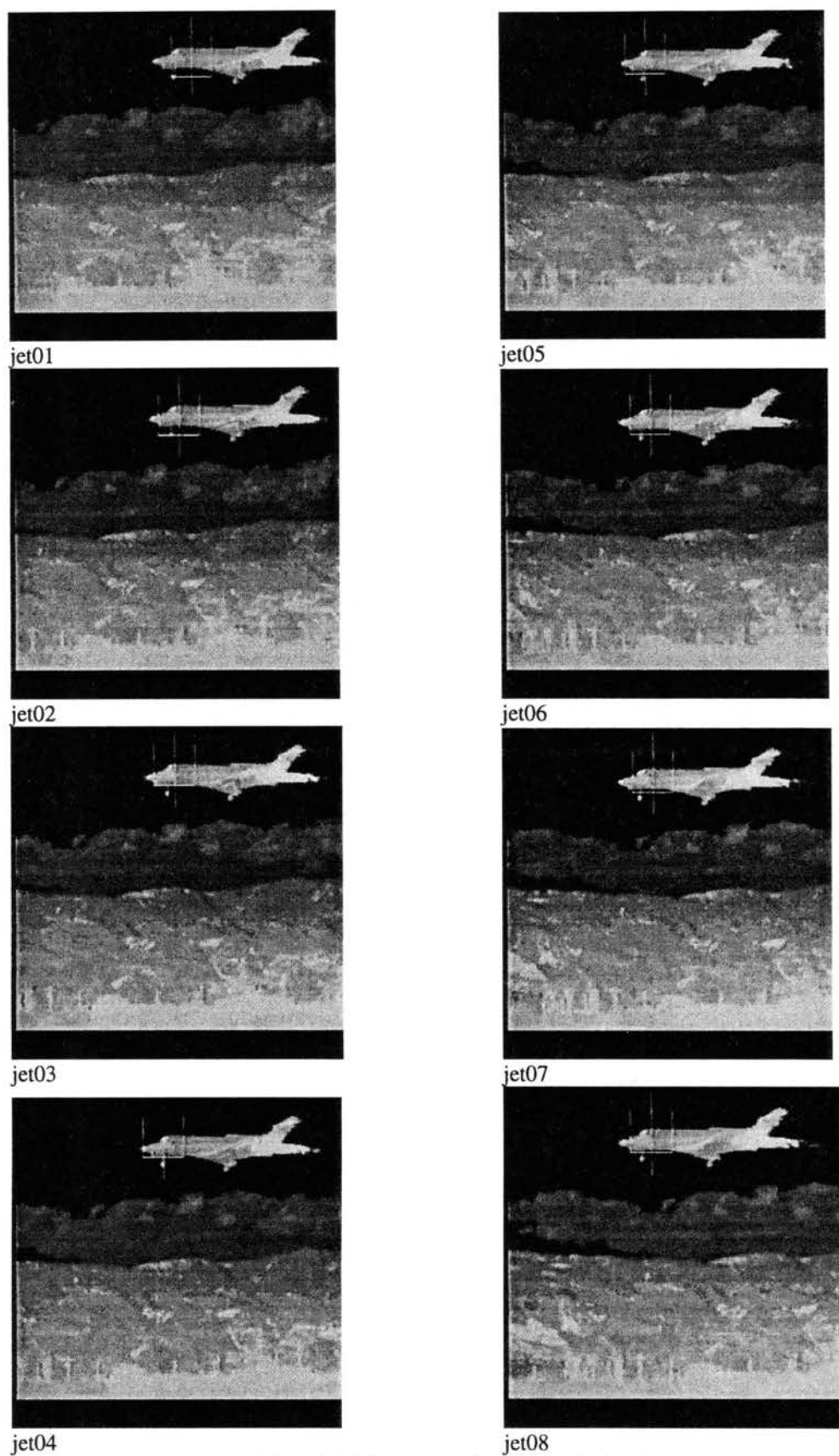the original image size (256x256)

jet01

jet05

jet02

jet06

jet03

jet07

jet04

jet08

Fig. 5.2 The example of an original sequence

jet09

jet13

jet10

jet14

jet11

jet15

jet12

jet16

Fig. 5.2 The example of an original sequence

jet17

jet21

jet18

jet22
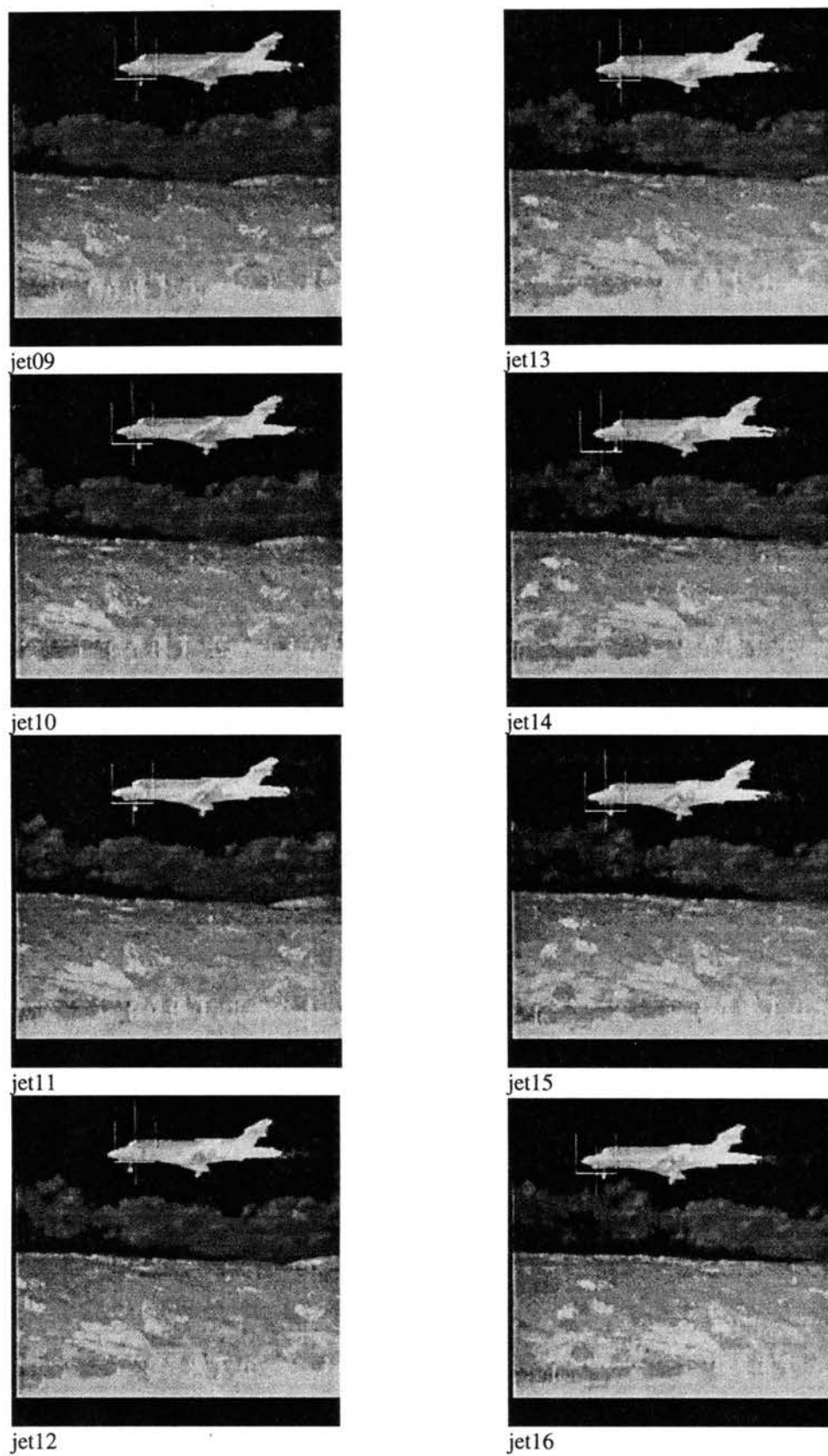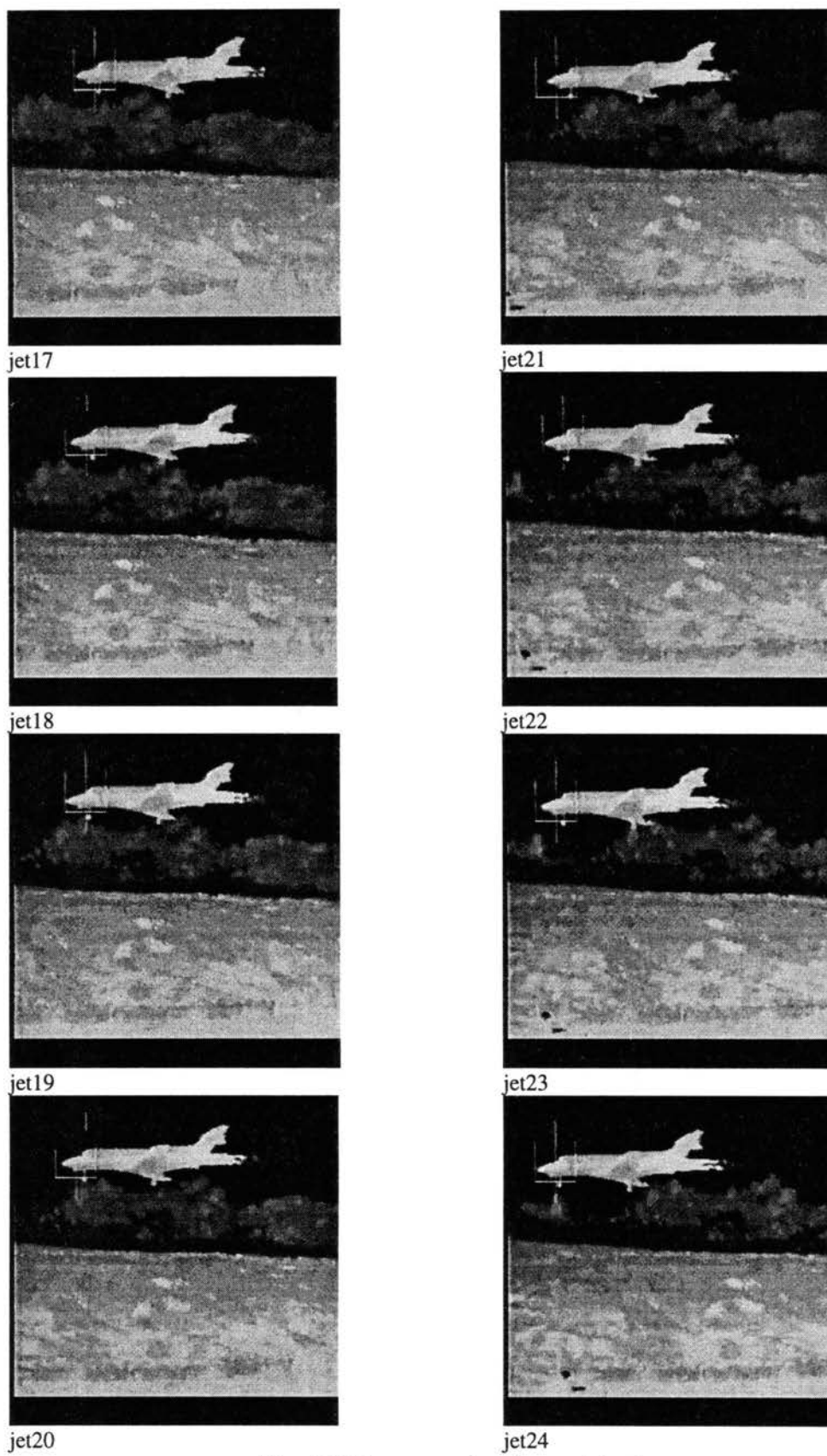
jet19

jet23

jet20

jet24

Fig. 5.2 The example of an original sequence

jet25

Fig. 5.2 The example of an original sequence
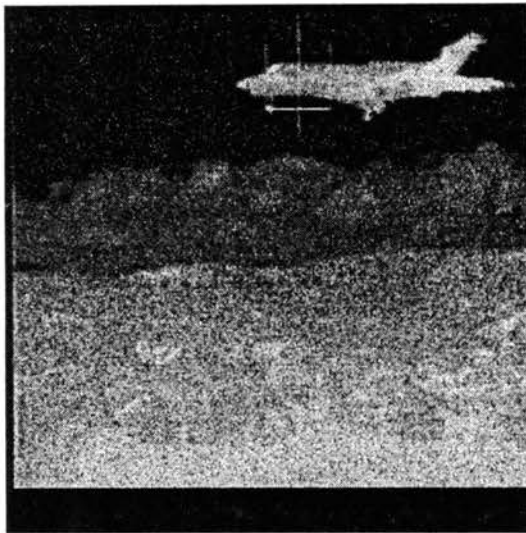
Original Image

MP Level 1 (128x128)

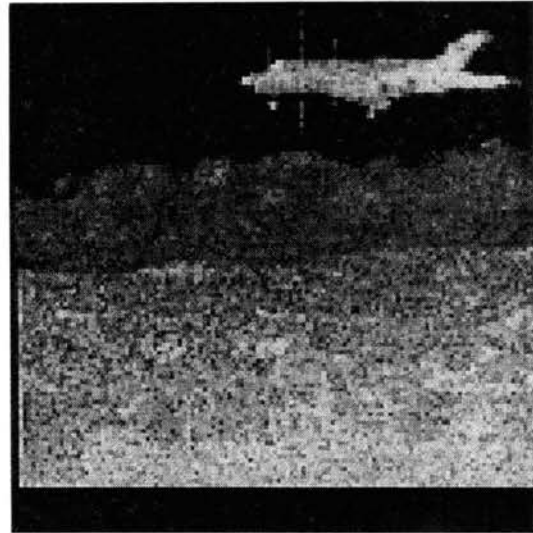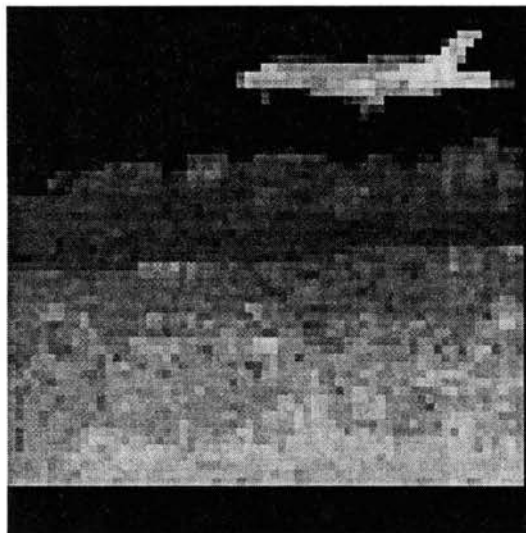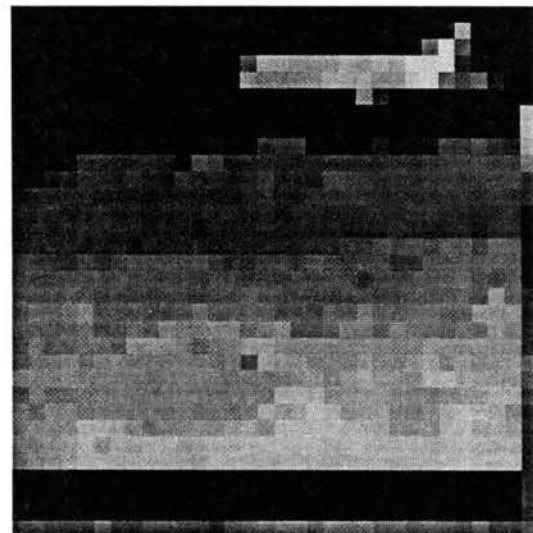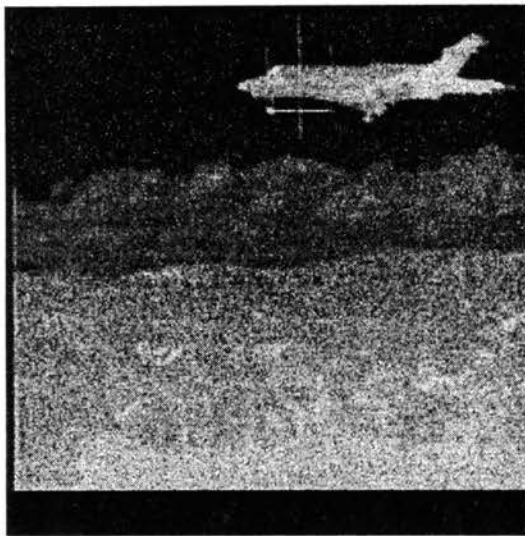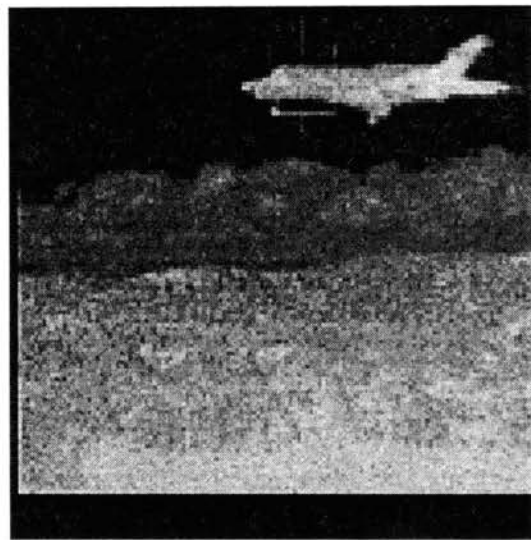MP Level 2 (64x64)

MP Level 3 (32x32)

Fig. 5.3(a) The MP for frame #1 of the corrupted sequence with the image size expanded

to the original image

Original Corrupted Image

GP level 1 (128x128)
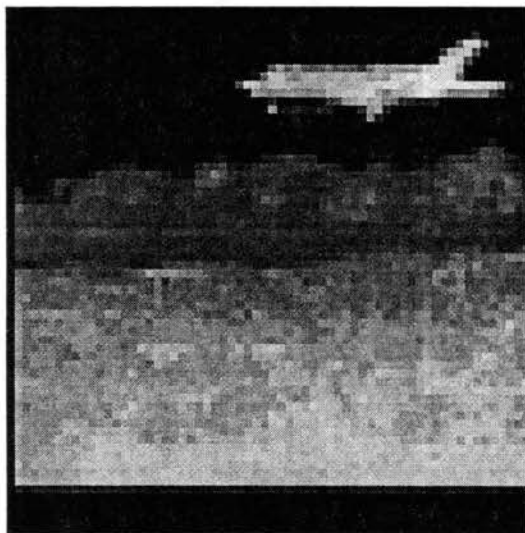
GP Level 2 (64x64)
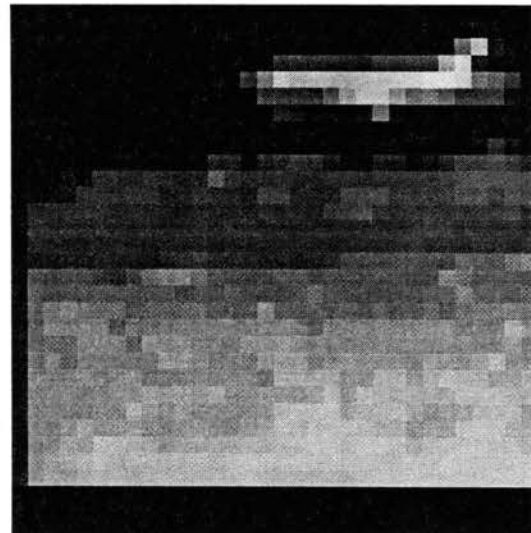
GP Level 3 (32x32)

Fig. 5.3(b) The GP for frame #1 of the corrupted sequence

window. It can be observed that the airplane is severely blurred and the internal information is lost in the third level of GP. On the other hand, the MP provides a more robust performance in terms of elimination of image noise and insignificant detail as well as preserving the integrity of the target. Comparing Fig. 5.1(b) (MP without noise) and Fig. 5.3(b) (MP with noise), it also demonstrates the robustness of the MP since the shapes of the target in those two figures in the third level are the same. Therefore, the MP can eliminate the noise and preserve the shape of the target compared with the GP.

To give a quantitative measure of the experiments performance, the mean square error (MSE) of GP and MP is calculated. The MSE criterion is defined by the following equation:

$$MSE = \frac{1}{M \times M} \sum_{m=1}^{M} \sum_{n=1}^{M} |I(m,n) - I'(m,n)|^2 \tag{3.17}$$

where

$M \times M$ is the size of the image.

$I(m,n)$ is the original image.

$I'(m,n)$ is the filtered image before sampling.

Table 5.1 shows the MSE at each level of the pyramid. It is shown in this table that nonlinear filters outperform Gaussian linear filter.

TABLE 5.1

MSE at different levels of the pyramids for the GP (Fig. 5.3(b)) and the MP (Fig. 5.3(a))

operation	MSE level 1	MSE level 2	MSE level 3
GP	3601.53	3958.81	4860.11
MP(Open-close)	3287.67	3667.57	4135.62

Coarse-to-fine Search Results

To implement a multi-resolution search, the root level (the coarsest pyramid level) for a given target must be identified. In the entire sequence shown in Fig. 5.2, the object of interest is an airplane whose minor axis of the largest part of the target is 16 pixels in width, and there are no other targets in the scene. Using (3.29-3.31) defines the root level of the sequence as the third resolution level in the pyramid.

Applying the coarse-to-fine search technique to this sequence and using the third level of the pyramid as the root level, the object identification is performed using a binary-based coarse-to-fine template matching routine. The computation time of both the fixed resolution and the multi-resolution case are summarized in Table 5.2. For the fixed resolution case, the algorithm required approximately 846 seconds per frame on a Sun Sparc 20 computer. The multi-resolution structure needs approximately 5 seconds per frame. The results show a system performance improvement of 169 times the traditional fixed resolution method. Fig. 5.4 compares the computational requirements of the fixed resolution and the multi-resolution approaches.

Table 5.2

Computation Time for the Fixed Resolution, the GP and the MP

Pyramid Type	None	GP	MP
Time(Second)	846	5	5
Normalized Time	170	1.0	1.0

Computational Time
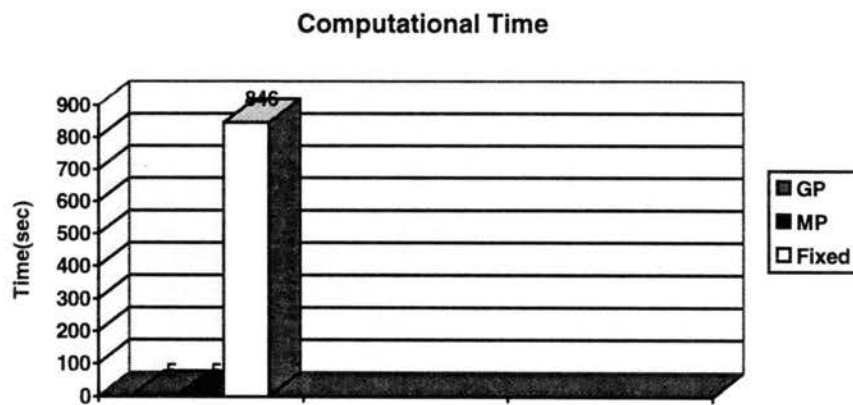
Fig. 5.4 Computational time of the fixed resolution and the multi-resolution approaches

Besides providing computational efficiency, morphological pyramid structure also increases system robustness. Using the binary coarse-to-fine template matching routine, the pixel localization errors for the original sequence and for the corrupted sequence with Gaussian distributed noise (SNR=9.7dB) are computed for both the fixed resolution and multi-resolution structures. The localization error is defined as the Euclidean distance between the detected position and the ground truth. These results are shown in Table 5.3 and Table 5.4, respectively.

Table 5.3

Detection Localization Error for an Original Sequence

Pyramid Type	None	GP	MP
Image 1	2.0	0	2.0
Image 2	1.0	1.0	1.0
Image 3	1.0	1.0	1.0
Image 4	2.2	1.0	1.0
Image 5	0	1.0	0.0
Image 6	0	5.0	0.0
Image 7	0.5	8.0	0.5
Image 8	2.2	3.2	3.1
Image 9	1.1	0.5	0.5
Image 10	1.1	1.1	1.1
Image 11	1.1	1.5	1.5
Image 12	2.5	2.5	2.5
Image 13	1.0	3.2	3.1
Image 14	2.5	2.5	2.5
Image 15	2.5	72.5	1.8
Image 16	3.3	74.6	3.3
Image 17	4.7	77.4	4.7
Image 18	4.7	91	5.6
Image 19	4.7	100.2	4.7
Image 20	3.9	103.0	4.7
Image 21	3.9	10.1	3.9
Image 22	3.2	6.5	4.6
Image 23	6.1	6.1	6.1
Image 24	4.6	54.6	4.0
Image 25	5.3	52.3	5.3
Average	2.6	27.2	2.7

Table 5.4

Detection Localization Error for the Noisy sequence (SNR=9.7dB)

Pyramid Type	None	GP	MP
Image 1	2.0	2.0	2.0
Image 2	1.0	6.1	2.1
Image 3	1.0	1.0	1.0
Image 4	2.2	1.0	1.0
Image 5	0	1.0	0
Image 6	0	4.1	0
Image 7	0.5	7.0	0.5
Image 8	2.2	3.2	3.2
Image 9	1.1	1.1	0.5
Image 10	1.1	1.1	1.1
Image 11	1.5	1.5	1.1
Image 12	2.5	2.5	4.3
Image 13	1.4	3.2	3.2
Image 14	2.5	2.5	2.5
Image 15	3.4	46.6	2.5
Image 16	3.4	83.3	3.4
Image 17	4.7	71.4	4.7
Image 18	4.7	89.0	5.6
Image 19	4.7	103.5	4.7
Image 20	3.9	69.8	4.7
Image 21	3.9	109.6	3.9
Image 22	3.2	3.2	4.6
Image 23	6.1	6.1	6.1
Image 24	5.3	56.3	5.3
Image 25	5.3	11.0	5.3
Average	2.7	27.5	2.9

Table 5.3 shows that the pixel localization error of the fixed resolution approach and the MP algorithm are similar, but the multi-resolution coarse-to-fine search time is considerably less than that of the fixed resolution method as shown in Table 5.2. To achieve the similar accuracy, the MP has a significant advantage in computational efficiency over the traditional fixed resolution approach. Comparing two multi-resolution approaches, MP gives localization error of 2.7 pixels and the GP localization error is 27.2 pixels. In some frames, the target is lost by the GP. For example, frame 14, 15, 19 and 20. In contrast, the MP located all the targets with much higher accuracy.

As can be seen from Table 5.4, the pixel localization error of the MP technique increases slightly in the presence of noise, compared with the uncorrupted sequences, but the algorithm is still capable of estimating the target location. The GP method is unable to locate the object during 8 frames of the sequence. The ability to find an object in a high clutter background allows the MP based identification system to provide a smaller pixel localization error with mean error of 2.9 pixels compared to 27.5 pixels of the GP system. Note that the MP-based approach achieves similar accuracy compared with the traditional fixed resolution approach, but with significantly less computational time than the fixed resolution method.

Fig. 5.5 plots the ground truth target positions, the positions detected by the MP with the original sequence and the corrupted sequence with SNR 9.7dB, respectively. It demonstrates that the positions detected by the MP, with or without noise, are fairly close to the ground truth. Therefore, it supports the fact that the MP approach is resilient to severe image corruption and clutter.

Fig. 5.6 compares the performance of the MP and the GP for the corrupted sequence with SNR 9.7dB. As can be seen from the graph, the MP results are close to the ground truth. In contrast, some of GP-generated points are scattered randomly because of target loss.
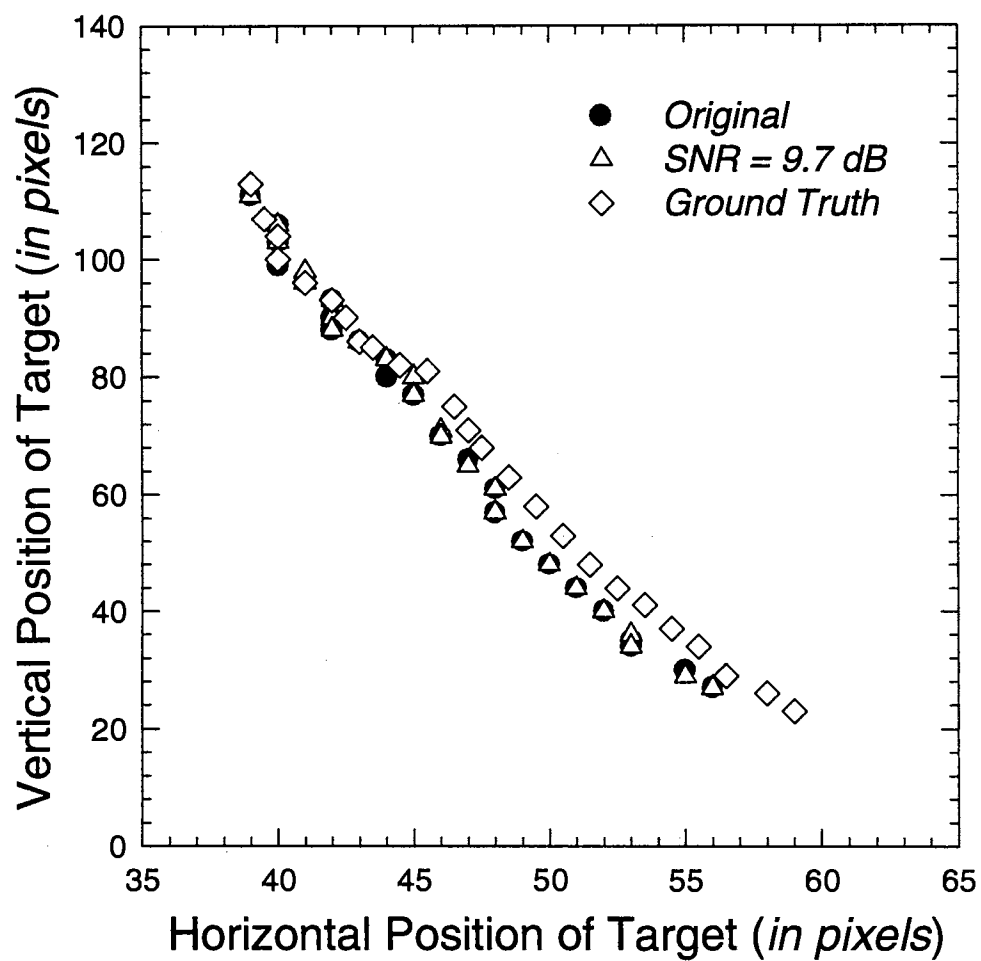
Fig. 5.5 Comparison of the target positions detected by the MP-based method for the

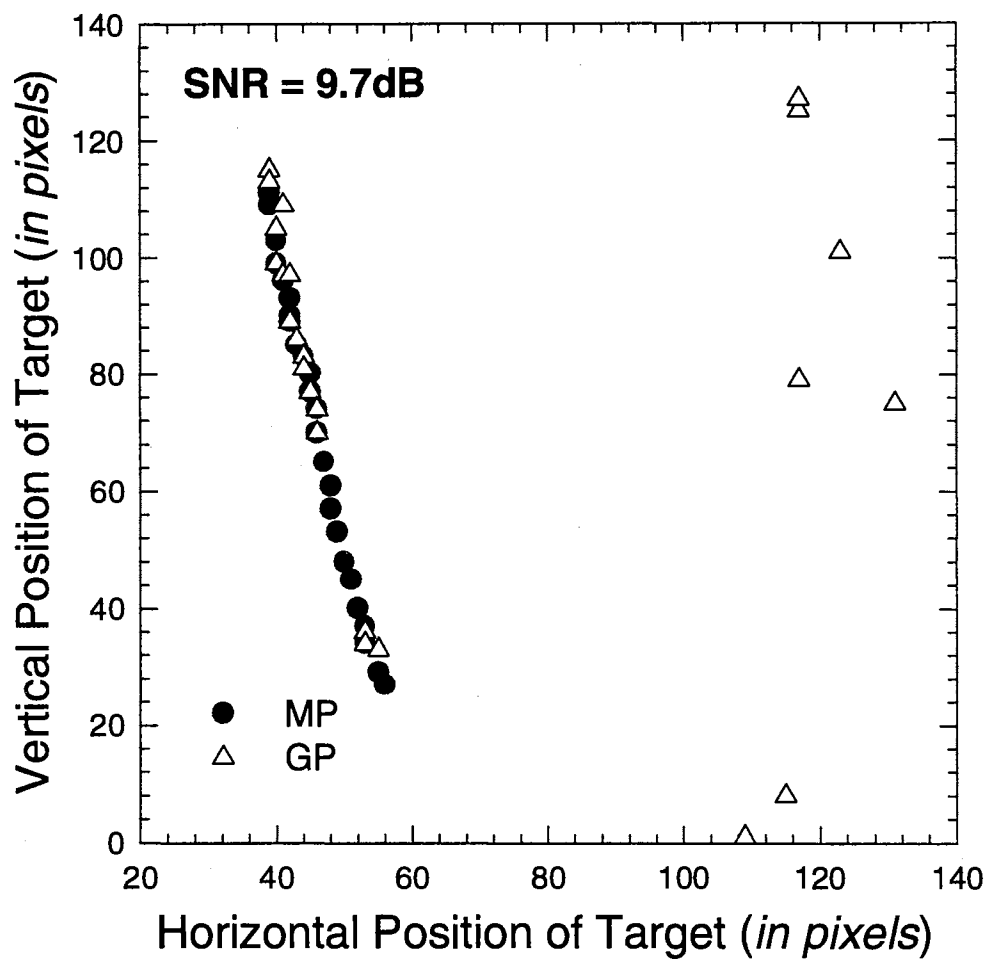original sequence and the corrupted sequence with the ground truth

Fig. 5.6 Comparison of the target positions detected by the MP with the GP for the

corrupted sequence with SNR 9.7dB

Selection of Parameters for Kalman Predictor

Chapter IV discussed integrating the MP detector with a Kalman predictor to enhance the MP-based search system. Although implementation of the Kalman filter is straightforward, the selection of the parameters for the Kalman filter can influence the accuracy of the tracking results.

Assuming that the velocity of the target is constant with a random drift, the α-β model can be used. The first step is to calculate the initial predicted state vector covariance matrix $P_{1|0}$ using equation (4.27). The parameters used for calculating $P_{1|0}$ are as follow:

variance of the initial i position $\sigma_i^2 = 30.0$

variance of the initial j position $\sigma_j^2 = 1.0$

variance of the velocity drift in i direction $\sigma_{v_i}^2 = 0.1$

variance of the velocity drift in j direction $\sigma_{v_j}^2 = 0.5$

variance of the noise in i direction $\sigma_{n_j}^2 = 1$

variance of the noise in j direction $\sigma_{n_{jj}}^2 = 1$

variance of the velocity drift noise in i direction $\sigma_{u_j}^2 = 0.85$

variance of the velocity drift noise in j direction $\sigma_{u_j}^2 = 0.05$

Then, (4.20) and (4.21) are used to obtain and store the gains (α_t and β_t) for each t. Next, the target position obtained by the MP detector is given to the Kalman filter. The

filtered position $\hat{i}_{t|t}$ is then computed using (4.15), and the predicted velocity $\hat{v}^i_{t+1|t}$ is

obtained by using (4.17). Finally, (4.14) is utilized to acquire the predicted position $\hat{i}_{t+1|t}$.

(this is repeated for the corresponding j terms).

Predicted Results vs. the Ground Truth

Table 5.5 shows the results of the Kalman predicted localization errors of the

original sequence and the corrupted sequence. The localization error is defined as the

Euclidean distance between the ground truth and the predicted position. The results, both

for the original and the corrupted case, are obtained by the Kalman filter with the

parameters presented in the last section. The observation data is provided by the MP-

based identification method for both cases. It can be seen that the localization errors for

both conditions are fairly low, only 3 pixels.

The comparison of the predicted target positions for the original sequence and the

corrupted sequence (SNR=9.7db) with the ground truth is shown in Fig. 5.7. It can be

seen that the predictor is effective under both conditions.

Table 5.5

Prediction Localization Error

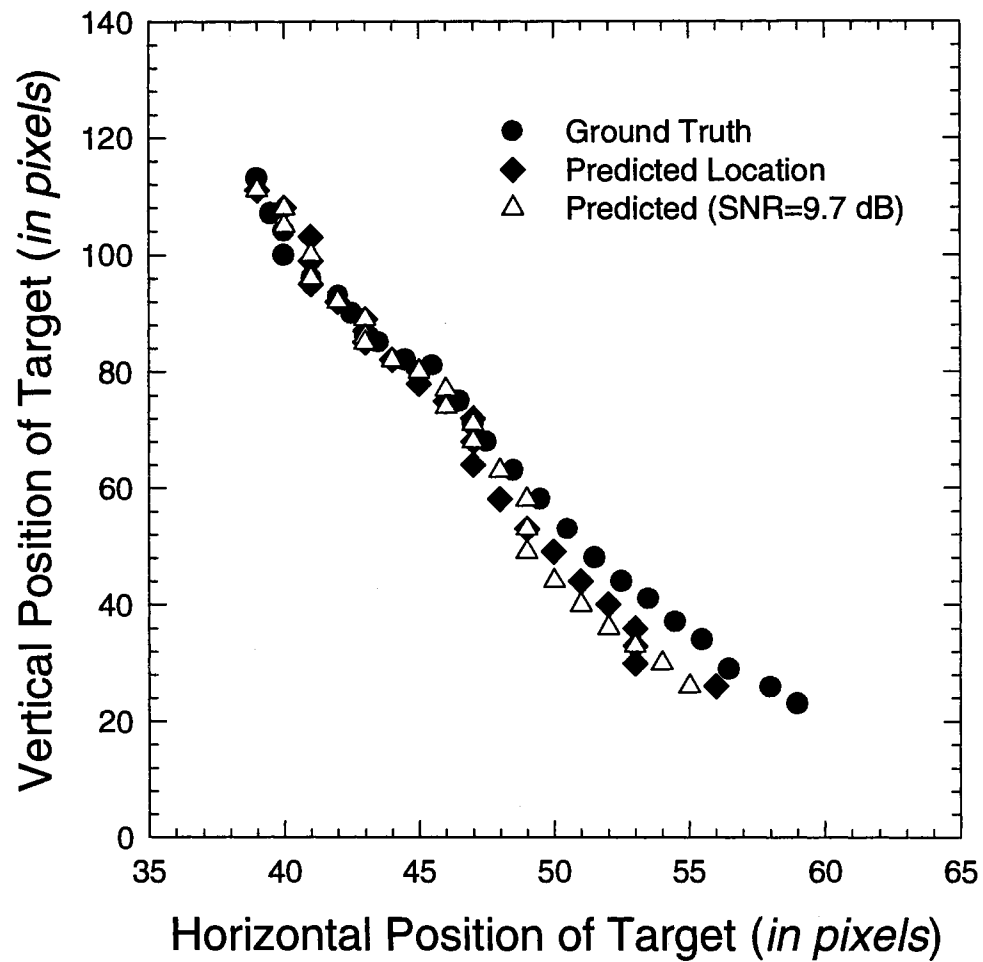|	original	Noisy(SNR=9.7dB)
Image 1	2.0	2.0
Image 2	1.0	1.0
Image 3	1.0	1.4
Image 4	1.0	1.4
Image 5	0	1.0
Image 6	1.0	1.0
Image 7	1.1	1.1
Image 8	0	1.0
Image 9	0.5	0.5
Image 10	0.5	0.5
Image 11	1.1	1.1
Image 12	2.0	3.4
Image 13	3.1	4.1
Image 14	3.0	4.0
Image 15	5.2	5.2
Image 16	5.2	6.5
Image 17	5.2	5.6
Image 18	5.6	5.6
Image 19	6.1	5.6
Image 20	4.6	3.9
Image 21	4.6	3.9
Image 22	4.0	3.2
Image 23	5.3	5.3
Image 24	6.0	6.8
Image 25	5.4	6.3
Average	3.0	3.3

Fig. 5.7 Comparison of the predicted target positions for the original sequence and the

corrupted sequence with the ground truth

Summary

The comparative results presented in this Chapter clearly demonstrate that the nonlinear morphological pyramid-based identification and tracking scheme outperforms the traditional fixed resolution approach and the linear Gaussian pyramid methods. The MP- based approach provides more accurate position information on the target even under severe noise and high clutter, as compared to the linear Gaussian pyramid, and is more computationally efficient compared to the fixed resolution approach. Comparing the ground truth with the Kalman-predicted results demonstrates the power of the predictive filter. With the right model and optimal parameters, the Kalman filter gives reliable predicted target positions under low signal-to-noise ratio conditions.

CHAPTER VI

CONCLUSIONS AND FUTURE WORK

In this chapter, the work presented in this research is summarized, the major contributions of this research are highlighted and several aspects for future work are suggested.

Summary and Conclusions

In this research, an automatic target recognition technique that utilizes the morphological pyramid is presented. With the pyramidal structure, potential targets are first identified from coarse resolution image representations, then identification and tracking of the target are performed in a small region of interest within higher resolution representations. The morphological pyramid allows the rapid identification of target shape and is resilient to severe image corruption and clutter. This new automatic target identification algorithm based on morphological pyramids allows efficient execution and performance enhancement. The nonlinear morphological pyramid preserves crucial features such as edge, target shapes and internal target intensity information. These features are utilized to reliably locate targets. The results obtained by the MP approach to

object identification clearly demonstrate the advantages of utilizing morphological pyramids as compared to those attained via a single resolution image or linear pyramids such as Gaussian pyramids. The results from this identification approach provide more accurate position of the objects in the scene. It should be observed that the MP based identification algorithm requires relatively simple computation and may be implemented on a parallel architecture.

Contributions

The following are major contributions of this research. A new automatic target recognition technique based on nonlinear morphological pyramids has been developed. Compared with linear Gaussian pyramid-based identification and tracking techniques, this MP based identification and tracking method has lower edge localization error without adding any more computational complexity. The MP-based identification and tracking technique developed in this research is able to locate an object two orders of magnitude faster than conventional fixed-resolution techniques while maintaining the similar localization errors.

This MP-based identification and tracking technique utilized edge-based coarse-to-fine search strategy to further reduce the computational cost. The traditional gradient-based edge detectors are sensitive to noise, and produce false edges which leads to false identification. Therefore, a new edge detector which is based on the morphological pyramid structure has been developed in this research (Chpater IV). The segmentation and resulting edge detection yielded by this method is particularly effective in the presence of

impulse noise and demonstrates superior solution quality over standard fixed resolution detectors and previous pyramidal approaches. This edge detector has numerous applications in feature-based image analysis.

The issue regarding the optimality of the structuring elements for generating the MP in identification and tracking has been thoroughly studied. The multiple SE has been proposed to generate the morphological pyramid for identification and tracking. The mathematical expression of the shape of the multiple SE has been first given in this research (Chapter III). The SE size has been defined which satisfies the morphological sampling condition.

The root level for initial identification has been analyzed (Chapter III). The optimal selection of the root level can reduce the computational cost of identification task as well as increase the accuracy of the target localization.

In the situation when the target or the camera is partially occluded by dirt or vegetation, a predictive filter is used. To enhance the identification and tracking quality further, the Kalman filter has been integrated with the MP-based identification and tracking technique.

<center>Future Work</center>

This research addresses many issues related to identification and tracking targets in high clutter and low signal to noise ratio conditions based on the morphological pyramid. There are several issues that need further investigation and refinement.

The MP-based identification and tracking system is able to find the target under high clutter and varying environment, but there are still open questions regarding the tracking and identification of multiple targets. The selection of the threshold for coarse-to-fine template matching needs to be further analyzed based on larger image sequences with different target types and background.

Although the algorithms developed in this research can be applied to process images from a variety of sensors such as forward looking infrared (FLIR), synthetic aperture radar (SAR) and video, the applications are focused on IR images. Effective use of multisensor information for target identification and tracking will result in more reliable ATR systems. For example, SAR sensors provide better penetration of smoke, dust and bad weather. Using SAR images for target identification when target is occluded by dust will provide better performance than using only IR images.

The knowledge-based recognition technique and approaches based on physical principles, detection theory, statistical techniques and neural networks could be investigated, since no single approach is likely to be the solution to all ATR problems.

5. REFERENCES

Acton, S. T., A. C. Bovik and M. M. Crawford, "Anisotropic diffusion pyramids for image segmentation," *Proceedings of the IEEE International Conference on Image Processing*, pp.13-16, Austin, Texas, 1994.

Acton, S. T., "A pyramidal edge detector based on anisotropic diffusion," *Proceedings of the IEEE International Conf. on Acoustics, Speech and Signal Processing*, Atlanta, Georgia, 1996.

Adelson, E. H., E. Simoncelli and R. Hingorani, "Orthogonal pyramid transforms for image coding," *Proceeding of the SPIE symposium on Visual Communications and Image Processing*, vol. 845, pp.50-58, 1987.

Ayala, I. L. *et al.*, "Moving target tracking using symbolic registration," *IEEE Trans. Pattern Anal. and Mach. Intell.*, pp. 515-520, 1982.

Bhanu, B., "Automatic target recognition: state of the art survey," *IEEE Trans. Aerosp. Electron. Syst.* vol. AES-22, pp364-379, 1986.

Burt, P. J., "Multiresolution techniques for image representation, analysis, and 'smart' transmission," *Proceedings of SPIE Symposium on Visual Communications and Image Processing*, vol. 1199, pp.2-15, 1989.

Burt, P. J., "Smart sensing within a pyramid vision machine", *Proceedings of the IEEE*, vol. 1, pp. 1006-1015, 1988.

Burt, P. J., T. Hong and A. Rosenfeld, "Segmentation and estimation of region properties through cooperative hierarchical computation," *IEEE trans. on Systems, Man, and cybernetics,* vol. 11, no. 12, 1981.

Burton, M. and C. Benning "Comparison of imaging infrared detection algorithms: Infrared technology for target detection and classification," *Proceedings of the SPIE,* vol. 302, pp. 26-32, 1981.

Cantoni, V. and S. Levialdi, *Pyramidal Architectures for Computer Vision,* Plenum Press, New York, 1994.

Chen, W. and S.T. Acton, "Processing issues in the automated inspection of power insulators," *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, pp.* 24-29, San Antonio, 1996.

Chen, W. and S.T. Acton, "Object identification and tracking using morphological pyramids," Submitted to *Journal of Electronic Imaging.*

Chen, W. and S. T. Acton, "An edge detector based on morphological pyramids," Submitted for Presentation at *the 1997 IEEE International Conference on Image Processing.*

Dorrough, D. C., *et al.,* "A multi-sensor multi-mode tracker approach to missile ship targeting," *Proceedings of the Tri-Service Wrokshop on Missile Ship Targeting,* 1982.

Dyer, C. R. and A. Rosenfeld, *Thinning Algorithms for Grayscale Pictures,* Computer Science Technical report Series, TR-610, DAAG53-76C-0138, University of Maryland, College Park, Md., 1977.

Florencio, D. A. F. and R. W. Schafer, "A non-expansive pyramidal morphological image coder," *Proceedings of IEEE international Conference on Image Processing*, Nov. 1994.

Florencio, D. A. F. and R. W. Schafer, "Homotopy and critical morphological sampling," *Proceedings of SPIE Symposium on Visual Communications and Image Processing*, Chicago, Sept. 1994.

Fong, Yu-shan, C. A. Pomalaza-Raez and X. H. Wang, "Comparison study of nonlinear filters in image processing application," *optical Engineering*, Vol. 28, pp. 749, 1989.

Giardina, C. R. and E. R. Dougherty, *Morphological Methods in Image and Signal Processing*, Prentice Hall, New Jersey, 1987.

Goetcharian, V., "From binary to gray level tone image processing by using fuzzy logic concepts", *Pattern Recognit.* 12, 1980, 7-15.

Goetcharian, V., *Parallel Image Processes and Real-time Texture Analysis*, Doctoral thesis, University College, London, 1979.

Hadwiger, H., *Vorlesungen uber inhalt, oberflache, und isoperimetrie*, Springer-Verlag, Berlin, 1957.

Haralick, R. M. and L. G. Shapiro, *Computer and Robot Vision*, Addison-wesley, New York, 1993.

Haralick, R. M., X. Zhuang, C. Lim, and J. S. J. Lee, "The digital morphological sampling theorem," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2067-2090, Dec. 1989.

Heijmans, H. J. A. M. and A. Toet, "Morphological sampling," *CVGIP: Image Understanding*, vol. 54, pp. 384-400, Nov. 1991.

Holben, R. D.,"An MTI (moving target identification) algorithm for passive sensors," *Proceedings of the IEEE NAECON*, pp.114-121, 1980.

Kalman, R.E., "A new approach to linear filtering and prediction problem," *Journal of Basic Engineering*, pp.35-41, 1969.

Kirsch, R. A. et al., "Experiments in Processing Pictorial Information with a Digital Computer," *Procedings of the Eastern Joint Computer Conference*, 1957, pp. 221-229.

Maragos, P, "Differential morphological operators for image and video compression," *IEEE Trans. on Image Processing*, Vol. 5, pp. 922-937, 1997.

Maragos, P. and R. W. Schafer, "Morphological filters-- Part II: their relations to median, order-statistic, and Stack filters," *IEEE Trans. on ASSP*, Vol. 35, pp.1170-1184, 1987.

Matheron, G., *Random Sets and Integral Geometry*, Wiley, New York, 1975.

Meyer, F., "Contrast feature extraction", *Quantitative Analysis of Microstructures in Material Sciences, Biology, and Medicine*, Special Issue of Practical Metallography, Stuttgart, 1977.

Milgram, D.L. and A. Rosenfeld, "Algorithms and hardware technology for image recognition," *Final report to U. S. Army Night Vision and Electro-Optics Lab., Fort Belvoir, Va., Mar. 31, 1978.*

Miller, J.L. *Principles of Infrared Technology*, Reinhold, New York, 1993.

Minor, L. G., and J. Sklansky, "Detection and segmentation of blobs in infrared images," *IEEE Trans. on Systems, Man, and cybernectics,* pp216-232, 1981.

Morales, A., R. Acharya, and S. Ko, "Morphological pyramids with alternating sequential filters," *IEEE Trans. Image Processing,* vol. 4, No. 7, pp. 965-977, 1995.

Salembier, P. *et al.* "Morphological operators for image and video compression," *IEEE Trans. on Image Processing,* Vol. 5, pp. 881-898, 1997.

Politopoulos, A.S., " An algorithm for the extraction of target-like objects in cluttered FLIR imagery," *IEEE Aerospace and Electronic Systems Society Newsletter,* pp. 23-27, 1980.

Pratt, W. K., *Digital Image Processing,* John Wiley & Sons, New York, 1991.

Ramesh, V. and R. M. Haralick, "Performance characterization of edge detectors," *Applications of Artificial Intelligence X: Machine Vision and Robotics,* Vol. 1708, pp. 252-266, 1992.

Rosenfeld, A., "Connectivity in digital pictures", *J. Assoc. Comput. Mach.* 17, No.1, 1970, 146-160.

Sage, A. P. and J. L. Melsa, *Estimation Therory with Applications to Communication and Control,* McGraw-Hill, New York, 1982.

Schneier, M., "Using pyramids to define local thresholds for blob detection," *IEEE Trans. Pattern Anal. and Mach. Intell.,* Vol. PAMI-5, pp345-349, 1983.

Schonfeld, D. and J. Goutsias, "Optimal morphological pattern restoration from noisy binary images," *IEEE trans. Patt. Anal. Machine Intell.,* vol. 13. no. 1, Jan. 1991.

Serra, J., "An Introduction to Mathematical Morphology," *Computer Vision, Graphics, and Image Processing,* vol. 35, pp283-305, 1986.

Serra, J., *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

Sidiropoulos, N. D. *et al.*, "MAP signal estimation in noisy sequences of morphologically smooth images," *IEEE Trans. on Image Processing*, Vol. 5, pp. 1088-1093, 1997.

Song, J. and E. J. Delp, "The analysis of morphological filters with multiple structuring elements," *Comput. Vision Graphics Image Process.* vol. 50, pp. 308-328, 1990.

Sternberg, S. R., "Grayscale morphology," *Comput. Vision Graphics Image Process.* vol. 35, No. 3, pp. 333-355, 1986

Sun, F. K. and P. Maragos, "Experiments on image compression using morphological pyramids," *Proceedings of SPIE Symposium on Visual Communications and Image Processing*, 1989.

Toet, A.,"A morphological pyramidal image decomposition," *Pattern Recognit. Lett.* 9, pp. 255-261, May 1989.

Walker, P. F., " Smart weapons in naval warfare," *Scientific American,* vol. 248, pp.53-61, 1983.

Wells, W. M., "Efficient synthesis of Gaussian filters by cascaded uniform filters," *IEEE Trans. on Pattern Anal. and Mach. Intell.,* Vol. PAMI-8, no. 2, pp15-26, 1986.

Uhr, L., "Highly parallel, hierarchical, recognition cone perceptual structures," *Parallel Computer Vision*, Academic Press, London, pp. 249-287, 1987.

VITA

Wei Chen

Candidate for the Degree of

Doctor of Philosophy

Thesis: MORPHOLOGICAL IMAGE PYRAMIDS FOR AUTOMATIC TARGET RECOGNITION

Major Field: Electrical Engineering

Biographical:

Education: Received Bachelor of Science degree in Physics from Nanjing University, Nanjing, China in July 1982; received Master of Science Degree in Electrical and Computer Engineering from Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China in July 1988; Completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May, 1997.

Professional Experience: Research Associate, Shanghai Institute of Technical Physics, from July, 1988 to May, 1991; Graduate Research Assistant, School of Electrical and Computer Engineering, Oklahoma State University, January, 1994, to present.