REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	15 Nov 96	Final Progress 26 May 93 - 30 Sep 96

4. TITLE AND SUBTITLE

Locating and Tracking Multiple Targets in High Resolution Imagery

5. FUNDING NUMBERS

DAAH04-93-G-0232

6. AUTHOR(S)

Gerald M. Flachs and Jay B. Jordan (PIs)
Zhonghao Bao and Charles Hardin

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)

New Mexico State University
Klipsch School of Electrical & Computer Engineering
P. O. Box 30001, Dept. 3-0
Las Cruces, NM 88003-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

31055.1-EL-H

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Mathematical foundations are developed for robust multiple target tracking systems. A flying spot locator and a meaningful trajectory state machine are presented that utilizes visual and trajectory information to assist in locating, recognizing and tracking multiple targets in complex scenes. Three dimensional multiple target trajectories are established using multiple view sensors. The reliability and real-time nature of the algorithms provide a sound basis to update and extend the performance of current real-time video tracking systems to more complicated missions with multiple targets. Algorithms for extracting multiple 3D target trajectories from multiple view high resolution video cameras is of direct value to Army range instrumentation programs.

14. SUBJECT TERMS	15. NUMBER IF PAGES
Research, Engineering, Electronic Vision	40
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Locating and Tracking Multiple Targets in High Resolution Imagery

Final Progress Report

Authors:

Dr. Jay B. Jordan

Dr. Gerald M. Flachs

Dr. Zhonghao Bao

Mr. Charles Hardin

November 15, 1996

U. S. Army Research Office
Grant: DAAH04-93-G-0232

Department of Electrical and Computer Engineering
New Mexico State University
Las Cruces, New Mexico 88003

TABLE OF CONTENTS

1. STATEMENT OF RESEARCH PROBLEM

Significant research and development effort has been devoted to creating smart weapon systems capable of detecting air and ground targets in widely varying hostile environments. These systems use the most modern multispectral sensor technology (infrared, visible, ladar, millimeter wave) to perform in complex scenes with widely varying environmental and countermeasure conditions. At White Sands Missile Range and other U.S. Army test facilities, optical tracking mounts equipped with high speed film cameras are used to record the trajectories of objects in multiple target missions for testing and evaluating modern weapon systems. The film is manually read by operators who locate and define the orientation of the targets. Information from different tracking stations is combined to establish accurate three dimensional trajectory profiles. The film reading task is tedious, expensive and time consuming due the large number of frames and the size and contrast of the target signatures. In many systems a carrier vehicle releases submunitions that are individually targeted. The trajectories of these submunitions is often the main point of interest. When the munitions are released, they are often obscured by smoke and other similarly shaped objects used in their release. Finding and tracking these submunitions is often a difficult task even for the human visual system.

The research sponsored by this grant is focused on the automation of the labor intensive task of precisely finding and tracking multiple targets to greatly reduce the time and costs involved in high resolution film data reduction. The specific objective of the research is to assist the Army in developing a computer vision theory and algorithms required for a high resolution video system capable of automatically reading multiple target and multiple view film data to obtain accurate three dimensional target trajectories.

2. SUMMARY OF RESULTS

Much of the theoretical background for this research [9-16] was established by the multisensor image analysis research funded by the U. S. Army Research Office (ARO) under grant DAAL03-87-K-0106. The background research focused on the development of mathematical foundations and methodologies for designing intelligent multisensor vision systems. These multisensor tracking methods are modified and extended to multiple target scenarios and applied to the high resolution film reading problem.

Reliable, real-time algorithms are developed for extracting three dimensional multiple target trajectories from multiple view high resolution digitized film data. Objects are located and tracked using both their visual characteristics and their trajectory information. Results indicate that meaningful

trajectory information allows multiple objects to be reliably tracked when the detector is operated with a large false alarm rate. Operating the detector with a high false alarm probability allows small and low-contrast objects to be located, identified, and tracked in highly cluttered backgrounds.

The reliability and real-time nature of the algorithms provide a sound basis to update and extend the performance of the current real-time video trackers to more complicated missions. Algorithms for extracting three dimensional multiple target trajectories from multiple view high resolution video cameras is of direct value to Army range instrumentation programs.

2.1 Mathematical Framework

A mathematical framework is developed for locating, tracking and forming three-dimensional trajectories for multiple targets using multiple tracking stations. A sensory scene model characterizes a scene as viewed by a sensor with many random factors contributed by atmospheric conditions, sensor noise and dynamic scene perturbations. A digitized image is considered as a composition of regions with different statistical characterizations. Some of the regions are natural scene components and others are the targets of interest. Statistical mixture separation methods are developed to separate the scene components. Targets are considered as scene components surrounded by other components or holes in scene components with given size, color and shape characteristics. Trajectory information is used to assist in identifying and tracking multiple targets using a priori knowledge of target dynamics.

2.1.1 Sensory Scene Model

Most electronic vision systems perceive their environment with sensors that make measurements in a multidimensional electromagnetic energy field conceptualized by Fig. 2.1. The observed values of these signals at any time are influenced by a myriad of factors in the energy field and the sensors. The overall effect of these factors makes it desirable to model the signals composing the scene as a stochastic or random process [1] in time, space and spectra. For multiple sensor systems, there is a dimension of the random process associated with each sensor. At any point in time, region of space and spectra, a multisensor system observes a single realization of the process.

In general, the joint distribution of the random variables composing the random process associated with the multisensor system is incredibly complex. However, in physical systems the random process describing the scene environment often contains small regions that are approximately wide sense stationary in space, time and spectra with relatively simple statistical properties [2]. The existence of such regions allows one to obtain information about the random process by examining spatially and

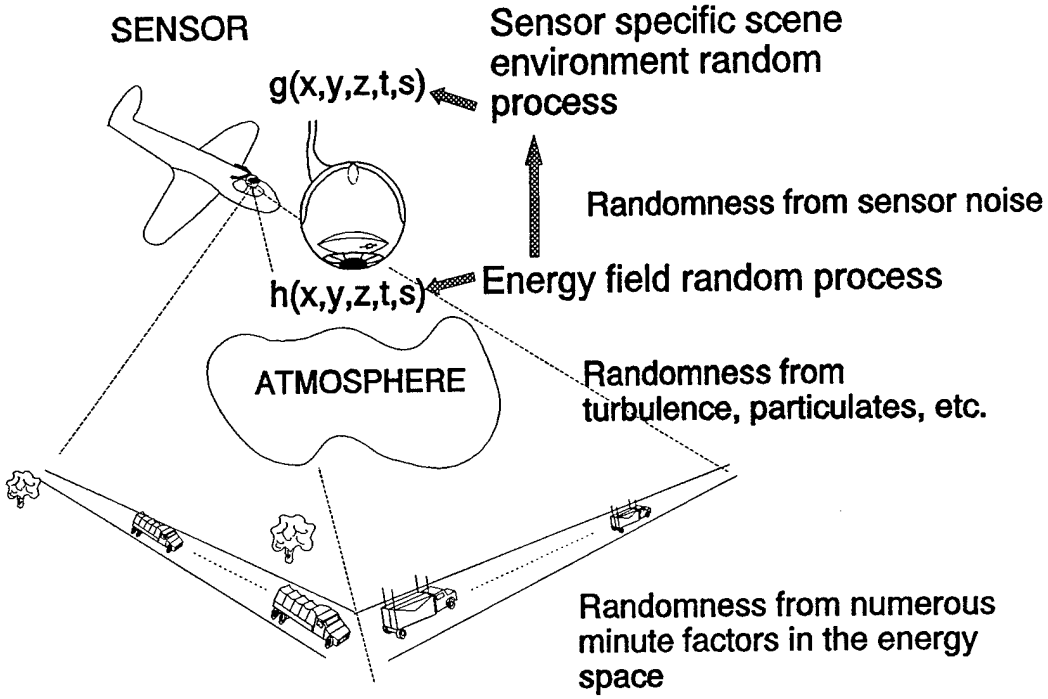


Fig. 2.1 Sensory Scene Model

temporally proximate samples. These properties provide the fundamental justification for many of the techniques presented.

The object/background dichotomy representation of a scene environment that arises in the context of many vision systems is an implicit statement of a priori information concerning the random process modeling the scene environment. Specifically, the a priori knowledge that there are two classes, objects and nonobjects, allows us to partition the set of random variables comprising the random process into two subsets described by two random processes; the process giving rise to observations in the object class, C_1, and the process giving rise to observations in the background class, C_2.

In terms of the random process model of the scene environment, the tasks of detecting the object in the background and separating it from the background are naturally posed as statistical hypothesis tests to determine which of the sets of random variables gave rise to the observed measurement(s). Formulation of the hypothesis test requires that the joint distributions of the random variables composing the objects and background be estimated [3].

The object recognition problem further partitions the set of random variables into several subsets

of random variables comprising different object classes. The problem of identifying or classifying an object is an m-ary hypothesis test to determine which of the object classes gave rise to the observed value(s). Again, formulation of this test requires the estimation of the joint distributions of the random variables composing each object class.

Although the statistical decision-making technique is conceptually simple, its application in practice can be quite difficult. The difficulty arises because in each scene, there is a continuum of random variables in the random process describing each class. It is not practical in real scenes to acquire enough information to fully characterize the complicated joint distributions of the random variables composing the random processes. For this reason, it is necessary to use some function (or set of functions) of the random variables that transforms the continuum to a finite set of random variables, $\{\mathbf{x}_1, \mathbf{x}_2,\ldots,\mathbf{x}_n\}$, called features. As the set is ordered, it is often denoted $\mathbf{X} = [\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n]$ where $\mathbf{X}$ is termed the feature vector. This reduction in dimensionality of the problem [4] is essential for the implementation of an actual system.

To illustrate the statistical decision-making process, consider the object locating and segmenting tasks where the random variables $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n$, are features common to both object regions and background. The set $x = \{x_1,x_2,..,x_n\} = [x_1,x_2,\ldots,x_n]$ is the observed feature vector, $f_{\mathbf{X}}(x_1,x_2,\ldots,x_n \mid C_1) = f_{\mathbf{X}}(x \mid C_1)$ is the conditional joint probability density or mass function (pdf) of the random variables under the object hypothesis, and $f_{\mathbf{X}}(x_1,x_2,\ldots,x_n \mid C_2) = f_{\mathbf{X}}(x \mid C_2)$ is the conditional joint pdf of the random variables under the background hypothesis.

The typical hypothesis test [5] based on observed values $x_1,x_2,\ldots,x_n$ is stated:

$H_0(C_1)$: $x_1,x_2,\ldots,x_n$ are observations from $f_{\mathbf{X}}(x_1,x_2,\ldots,x_n \mid C_1)$

$H_1(C_2)$: $x_1,x_2,\ldots,x_n$ are observations from $f_{\mathbf{X}}(x_1,x_2,\ldots,x_n \mid C_2)$

Decisions to classify the observations as being from object (H_0) or background (H_1) are typically made in such a manner that either a) the probability of making an error is minimized or b) the cost of making an error is minimized. A similar set of tests is constructed for the classification of objects into various object classes.

Elements of the feature vector $\mathbf{X}$ may come from transformations acting on measurements from a single sensor or acting on measurements from more than one sensor. Mathematically, the treatment of systems involving multiple sensors, multiple features, or multiple sensors and features is identical. The first step in object/background or object/object segmentation is determining the set of features that converts the raw sensor measurements into the feature measurements represented by $\mathbf{X}$. In practice, the scene from each sensor is sampled spatially and each sample is quantized into a finite number of gray-

levels. This creates a finite, but large feature vector. This space is reduced to a much smaller feature space by assuming that the distribution of gray level values in any small region of an image is an observation from a random process composed of simple mixture distributions. This leads to an effective statistic mixture model representation of an image.

2.1.2 Statistic Mixture Image Model

To find an efficient way to segment an image into its basic components, the distribution of a sensor's measurements over a complex scene is characterized as a mixture of component distributions that stem from the different background and target regions in an image. Due to the many random factors affecting a sensor's measurement, each of the component distributions often are well-behaved and have a Gaussian appearance. Therefore, image segmentation can be implemented [6,7,8] by separating a mixture into its components, locating the scene components and analyzing the shape and trajectory information associated with the components.

Consider an image to be the digitized output of a sensor, resulting from a mosaic of N x M picture elements from the scene. The value of each picture element or pixel in the digitized image is proportional to the average intensity of radiated or reflected energy from a small region in the scene. The digitized image is designated **X** and consists of M x N pixels $x(i,j) \in G=\{0,1,2,\ldots,B-1\}$ where B is the number of digitized color levels and $i \in \{0,1,2,\ldots, M-1\}$ and $j \in \{0,1,2,\ldots,N-1\}$ define the spatial coordinates. The probability that $x(i,j)$ takes on some color level $x \in G$, $(P(x(i,j)=x))$, is given by the probability density function $f(x)$ which is estimated by a normalized histogram $h(x)$, giving the relative frequency at which a gray level value x occurs in the digital image.

The quality of a relative frequency histogram $h(x)$ as an estimator of a multiple mode density function $f(x)$ generated by the scene components depends on many factors. The major factors include the number of samples from each mode (resolution) and the complexity of the distribution modes. The sensor pixel averaging and the many random factors perturbing the sensor measurement tend to enhance the Gaussian distribution assumption. Non-uniform lighting conditions, however, over large scene components often negates the Gaussian assumption. To retain the Gaussian assumption the size of a component region must be restricted in non-uniform lighting situations. Fortunately, target regions are often small regions and satisfy the Gaussian assumption.

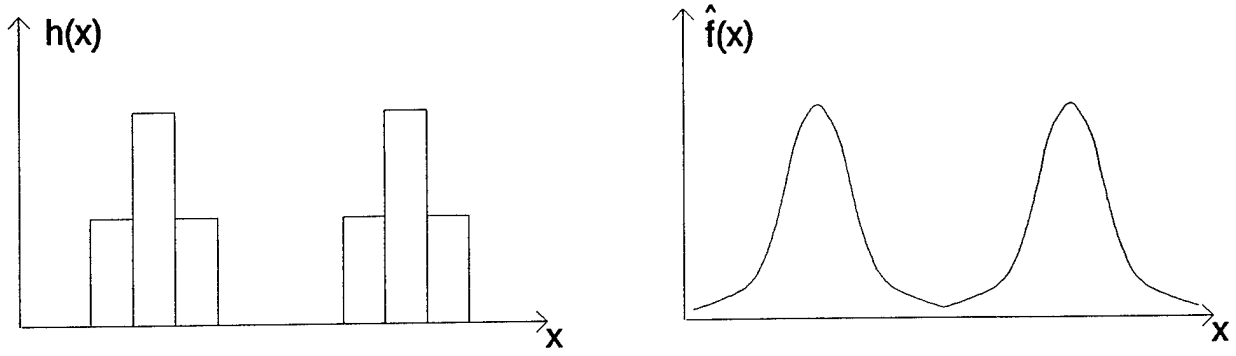
2.1.3 Identification of Mixture Components

Maximum likelihood estimation techniques [24] can be used to separate the Gaussian mixture

components. Consider a mixture distribution

$$f(X)=\sum_{i=1}^{k}\pi_i f_i(X)=\sum_{i=1}^{k}\frac{\pi_i}{(2\pi)^{\frac{m}{2}}|\Sigma_i|^{\frac{1}{2}}}e^{-\frac{1}{2}(X-M_i)^T\Sigma_i^{-1}(X-M_i)}, \quad (1)$$

where M_i is a mean vector, Σ_i is a covariance matrix, and π_i is a weight coefficient such that $\pi_1+\pi_2+\cdots+\pi_k=1$. When the number of components, k, is known, a maximum likelihood estimator can be used to estimate the parameters π_i, M_i, and Σ_i, i=1, 2, ..., k, which then can be used to separate the scene components.



(a) (b)

Fig. 2.2 A relative frequency histogram and its MLE function.

The maximum likelihood estimators of π_i, M_i, and Σ_i, i=1, 2, $\cdots$, k, are realized by maximizing $\prod_{j=1}^{N} f(X_j)$ with respect to π_i, M_i, and Σ_i under the constraint $\pi_1+\pi_2+\cdots+\pi_k=1$, where X_1, X_2, $\cdots$, and X_N, are random samples from the population with density function $f(X)$. For example if the relative histogram, h(x), of a sample shown in Fig. 2.2(a) is from a mixture of two normal components, the maximum likelihood component density functions would have the shape shown in Fig. 2.2(b). Theoretically, the maximum likelihood estimates are asymptotic to the true value of parameters π_i, M_i, and Σ_i.

To simplify the analysis, the logarithm of $\prod_{j=1}^{N} f(X_j)$ is taken since maximizing $\prod_{j=1}^{N} f(X_j)$ is equivalent to maximizing the logarithm of $\prod_{j=1}^{N} f(X_j)$. Hence, the criterion to be maximized is

$$J=\sum_{j=1}^{N}\ln f(X_j)-\lambda\left(\sum_{i=1}^{k}\pi_i-1\right)\ , \tag{2}$$

where λ is a Lagrange multiplier.

The derivative of J with respect to π_i is

$$\frac{\partial J}{\partial \pi_i}=\sum_{j=1}^{N}\frac{f_i(X_j)}{f(X_j)}-\lambda=\frac{1}{\pi_i}\sum_{j=1}^{N}q_i(X_j)-\lambda\ , \tag{3}$$

where

$$q_i(X)=\pi_i\frac{f_i(X)}{f(X)}\ . \tag{4}$$

Observe that

$$\sum_{i=1}^{k}q_i(X)=\frac{\sum_{i=1}^{k}\pi_i f_i(X)}{f(X)}=1\ . \tag{5}$$

Furthermore, if $\frac{\partial J}{\partial \pi_i}=0$,

$$\sum_{i=1}^{L}\pi_i\frac{\partial J}{\partial \pi_i}=\sum_{j=1}^{N}\left(\sum_{i=1}^{k}q_i(X_j)\right)-\left(\sum_{i=1}^{k}\pi_i\right)\lambda=N-\lambda=0, \tag{6}$$

and so

$$\lambda=N\ . \tag{7}$$

Hence, the solution of equation (3) is given by

$$\pi_i=\frac{1}{N}\sum_{j=1}^{N}q_i(X_j)\ . \tag{8}$$

The derivative of J with respect to M_i is

$$\frac{\partial J}{\partial M_i}=\sum_{j=1}^{N}\frac{\pi_i}{f(X_j)}\frac{\partial f_i(X_j)}{\partial M_i}=\sum_{j=1}^{N}q_i(X_j)\Sigma_i^{-1}(X_j-M_i)\ . \tag{9}$$

Since $\sum_{j=1}^{N} q_i(X_j) = N\pi_i$ by equation (8), the solution of equation (9) becomes

$$M_i = \frac{1}{N\pi_i} \sum_{j=1}^{N} q_i(X_j) X_j \ . \tag{10}$$

The derivative of J with respect to Σ_i ([21,22]) is

$$\begin{aligned} \frac{\partial J}{\partial \Sigma_i} &= \sum_{j=1}^{N} \frac{\pi_i}{f(X_j)} \frac{\partial f_i(X_j)}{\partial \Sigma_i} \\ &= \sum_{j=1}^{N} q_i(X_j) \{\Sigma_i^{-1}(X_j - M_i)(X_j - M_i)^T \Sigma_i^{-1} \\ &\quad - \Sigma_i^{-1} - \frac{1}{2} diag[\Sigma_i^{-1}(X_j - M_i)(X_j - M_i)^T \Sigma_i^{-1} - \Sigma_i^{-1}]\} \ , \end{aligned} \tag{11}$$

where *diag[A]* is a diagonal matrix, keeping only the diagonal terms of A. The solution of equation (11) is given by

$$\Sigma_i = \frac{1}{N_i} \sum_{j=1}^{N} q_i(X_j)(X_j - M_i)(X_j - M_i)^T \ . \tag{12}$$

Theoretically, the maximum likelihood estimate of π_i, M_i, and Σ_i, $i=1, 2, \ldots, k$, can be obtained by solving equations (8), (10), and (12), simultaneously. However, since $q_i(X)$ is a function of π_i, M_i, and Σ_i, for $i=1, 2, \ldots, k$, an iterative algorithm is used to obtain the MLE solution.

Iterative Solution [24] :

Step 1. Choose initial weights $\pi_i^{(0)}$, mean vectors $M_i^{(0)}$, and covariance matrices $\Sigma_i^{(0)}$, and compute $q_i^{(0)}(X_j)$ by

$$q_i^{(l)}(X_j) = \frac{\pi_i^{(l)} f_i^{(l)}(X_j)}{\sum_{i=1}^{k} \pi_i^{(l)} f_i^{(l)}(X_j)} \ , \tag{13}$$

where $f_i^{(l)}(X)$ is the normal density function with mean vector $M_i^{(l)}$ and covariance matrix $\Sigma_i^{(l)}$ in *l-th* iteration, for $i=1, 2, \cdots, k$ and $l=0, 1, \cdots$.

Step 2. Having calculated $\pi_i^{(l)}$, $M_i^{(l)}$, and $q_i^{(l)}(X_j)$ for the *l-th* iteration, compute $\pi_i^{(l+1)}$, $M_i^{(l+1)}$, $\Sigma_i^{(l+1)}$, and $q_i^{(l+1)}(X_j)$ for the *(l+1)-th* iteration by equations (8), (10), (12), and (13) respectively.

Step 3. If $q_i^{(l+1)}(X_j)$ and $q_i^{(l)}(X_j)$ are close enough for all $i=1, 2, \ldots, k$ and $j=1, 2, \ldots, N$, then stop. Otherwise, increase l by 1 and go to step 2.

Good initial values for the distribution parameters can be obtained from the relative frequency histogram h(x), by using the number of significant peaks in h(x) and estimates of their means and variances. With good estimates of these parameters only a few iterations are required to obtain good results. When the maximum likelihood estimator is applied and the parameters π_i, M_i, and Σ_i, $i=1, 2, \ldots, k$, of the component distributions are obtained, the scene components are obtain by using a Bayesian classifier [23] to segment the image into the regions corresponding to the mixture components. A pixel is assigned to class l if and only if

$$\pi_l f_l(X_j) \geq \pi_i f_i(X_j) , \qquad i=1,2,\ldots,k, \text{ and } i \neq l . \tag{14}$$

That is, an observation vector X_j is always assigned to the class with largest value of $\pi_i f_i(X_j)$. Fig. 2.3 shows how the Bayesian classifier works with a two component mixture. In the general case, when the expressions of $f_l(X_j)$ and $f_i(X_j)$ are plugged into inequality (14) and then taking the logarithm of the inequality, inequality (14) can be written as

$$\frac{1}{2}(X_j - M_l)^T \Sigma_l^{-1}(X_j - M_l) + \frac{1}{2}\ln|\Sigma_l| - \ln \pi_l \geq \frac{1}{2}(X_j - M_i)^T \Sigma_i^{-1}(X_j - M_i) + \frac{1}{2}\ln|\Sigma_i| - \ln \pi_i , \qquad i=1,2,\ldots,k, \text{ and } i \neq l . \tag{15}$$

Hence, X_j is assigned to class l if and only if inequality (15) is satisfied. Thus, the scene is partitioned into k regions corresponding to the mixture components.

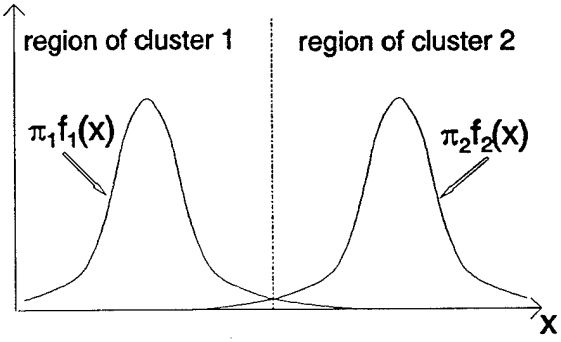


Fig. 2.3 Bayesian classification rule.

Gradient Mixture Separation

When mixture component distributions are not Gaussian, the maximum likelihood estimator may lead

to poor segmentation. For these situations a general-purpose parametric clustering algorithm is required to separate the mixture components without knowing the form of the component mixture distributions. A gradient mixture separation algorithm is presented for these situations. The basic idea is to use the natural peaks and valleys in the relative frequency histogram to separate the modes of the distribution. First, each peak in a frequency histogram is considered as a class and assigned a unique class symbol. The gradient is establish for the remaining feature vectors and each feature vector finds its own peak along the gradient and assumes the class identifier of the peak. The algorithm is stack implemented and it is much faster than the maximum likelihood estimator. The result of this operation is a classification matrix that assigns each feature vector to a mixture class. The classification matrix is used to separate the scene components.

2.1.4 Multiple Object Tracking Model

The basic objective in a multiple object tracking problem is to reliably and efficiently track multiple targets and establish their three dimensional trajectories using multiple cameras. From a single camera's perspective, objects are mapped into a plane and object trajectories can intersect in the plane. On a frame-to-frame basis, each tracking camera must locate the objects in it's field-of-view and correspond each object with a trajectory. This is a very difficult problem when the objects have similar shape and are occluded by smoke or other similarly shaped nontargets. Based on dynamic considerations, each object has a class of possible trajectories that are meaningful to the application. By eliminating the trajectories that are not meaningful to the application, the number of possible trajectory combinations is greatly reduced and the problem becomes tractable for real-time environments.

A meaningful trajectory is a motion of an object that is important to the application and satisfies all the dynamic constraints of the object in the given environment. For example, in a remote video surveillance system, purposeful motion to enter a secure area is meaningful, whereas random movement not directed toward the secure area is not. In most outdoor and some indoor environments, there is a significant amount of motion that is not meaningful. Efficient, reliable remote surveillance systems should not generate false alarms as a result of unimportant motion. An example of meaningful motion is an intruder walking or crawling toward an entryway or object, perhaps alternating quick dashes with hiding behind different objects.

The problem under consideration involves a missile or carrier vehicle that dispenses multiple submunitions. The carrier vehicle can be distinguished from the munitions based on shape and size parameters. The munitions, however, all have the same size and shape characteristics. The problem is

further complicated by erratic motion introduced by the human controlled camera pointing system. During the release of the submunitions, smoke and other objects used to release the targets are present. Since the submunitions and other objects all have similar shape and size, they are distinguished by their trajectories based on their aerodynamic characteristics. The concept of meaningful aerodynamic trajectories is used to solve the correspondence problem on a frame-to-frame basis, using a finite state machine to model the trajectory information. Two further difficulties arise in the tracking problem. First, the munitions can be occluded from view by noise or background features. The second problem results from false targets caused also by background noise and features. These two problems make the correspondence and tracking problems very challenging.

A finite state machine is used to model the trajectories under the conditions of occlusions and false targets. Each possible trajectory is associated with an initial state of a finite state machine. When two frames of data are available, a predictor is used to predict the locations for all possible meaningful trajectories. All trajectories within a certain region of possible motion are moved to a monitoring trajectory state. When trajectories attain five consecutive good predictions, they are moved to the valid trajectory state and the tracker locks on them. If an object is occluded in the valid trajectory state, then the predicted position is used to continue track but the confidence in the trajectory is lowered. If the trajectory confidence falls below a lower limit, the trajectory is terminated and must be rediscovered through the initial trajectory state logic. During the monitoring state, many false targets are often reported, but they are ignored until they attain five consecutive frames of valid trajectory. This process works well for the target tracking problem and it can be modified for the security protection problem. The sampling rate must be high enough to be able to track objects of interests based on their mobility. The concept of meaningful trajectory must also be modified to represent meaningful motion toward a secure area. This would require an additional alert state that trajectories would be placed after attaining the valid track state and making meaningful process toward the secure area. By combining the trajectories from several different viewing angles, three dimensional (3D) trajectories can be determined.

The input to the multiple target tracker comes from a target locator that identifies all possible target locations. Some of the objects found by the locator may be real targets while others may be false alarms. Furthermore, some objects may be occluded from the locator and are not reported as possible target locations. When the locator finds an object in a frame, the error in establishing the target position is generally very small.

The multiple object tracker consists of four modules: (1) a find module which continually attempts to find new trajectories, (2) a track module which selects objects to continue existing trajectories, (3) a

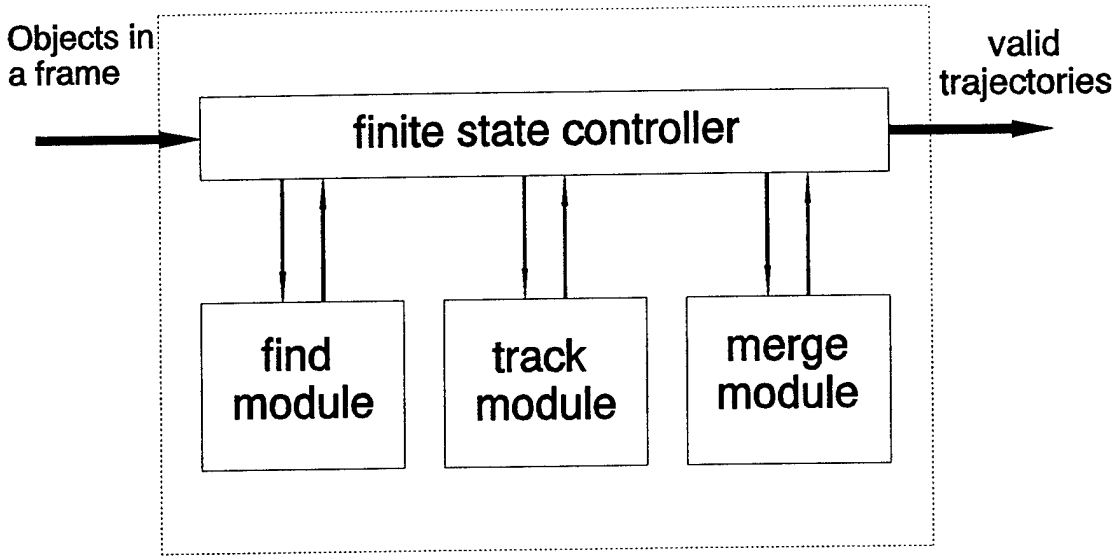


Fig. 2.4 The finite state tracking control structure.

merge module for merging trajectories, and (4) a finite state controller that selects the meaningful trajectories as shown in Fig. 2.4. When a new frame of possible targets arrives, the track module selects the best objects for advancing all valid trajectories. The find module attempts to establish new trajectories with all objects not selected by the track module. The merge module observers the trajectories to see if any should be merged into one. Based upon the results of these operations, the finite state controller adjusts all trajectory states, such as to create a trajectory, to terminate a trajectory, to increase a trajectory confidence, or to decrease a trajectory confidence.

The finite state controller is responsible for keeping track of all trajectories and establishing when they are meaningful, as shown in Fig. 2.5. When the find module establishes a new possible trajectory, the trajectory is assigned to the "Init" (initial) state. When the track module advances an existing trajectory a "Hit" signal is generated and the controller responds by adjusting the trajectory's confidence. Otherwise, a "!Hit" is generated and the controller will predict the location and lower the confidence of the trajectory. When the confidence in a trajectory falls below a given threshold it is terminated. When the merge module indicates to or more trajectories should be merged, merge signal is generated and the controller selects which trajectory is allow to continue and terminates the rest.

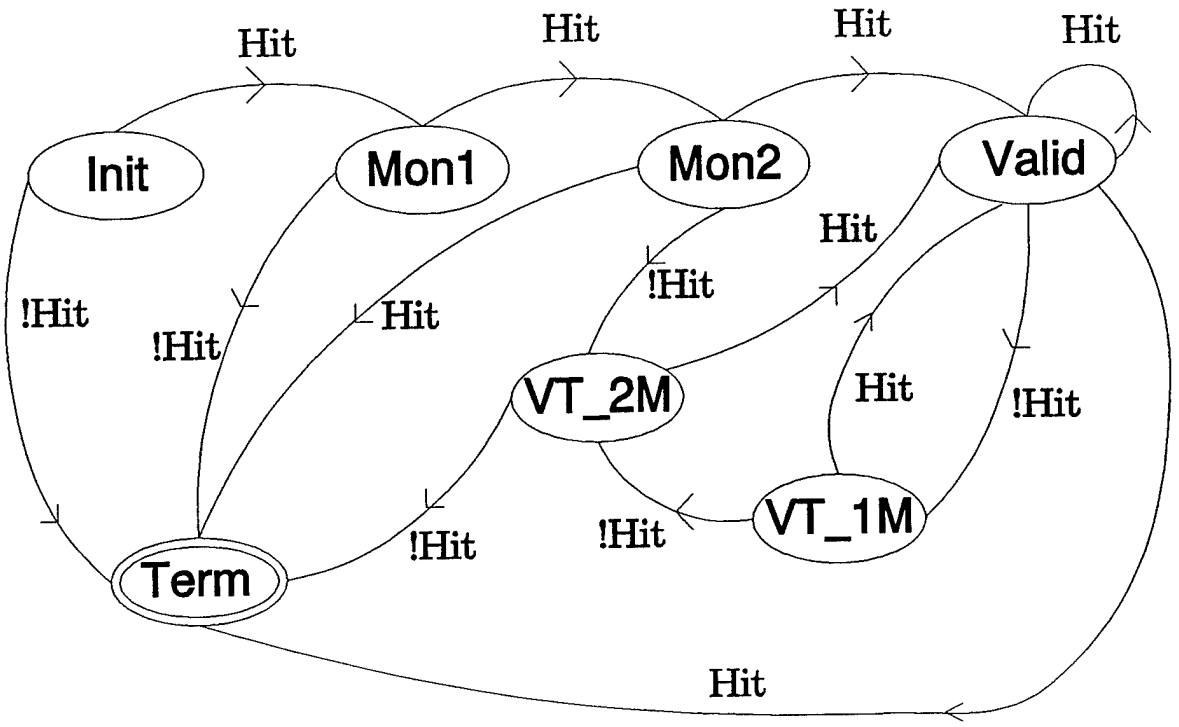
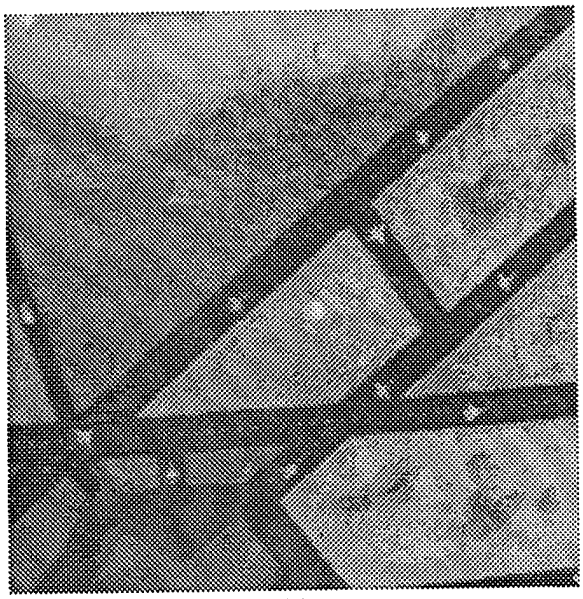


Fig. 2.5 The state diagram of a meaningful trajectory analyzer.

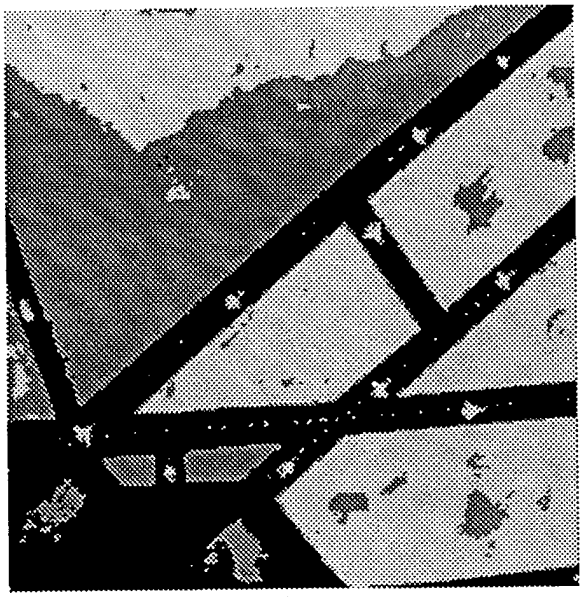
2.2 Mixture Separation and Segmentation Results

The mixture separation methods are applied to several problems to demonstrate their performance. The maximum likelihood mixture separation method works best when the mixture components are Gaussian and separated with multiple peaks. These conditions are often satisfied when the lighting conditions are uniform and targets are visible. The gradient mixture separation method is faster and works well when little is known about the distribution of the mixture components. Three problems are used to demonstrate and compare the performance of the mixture separation methods.

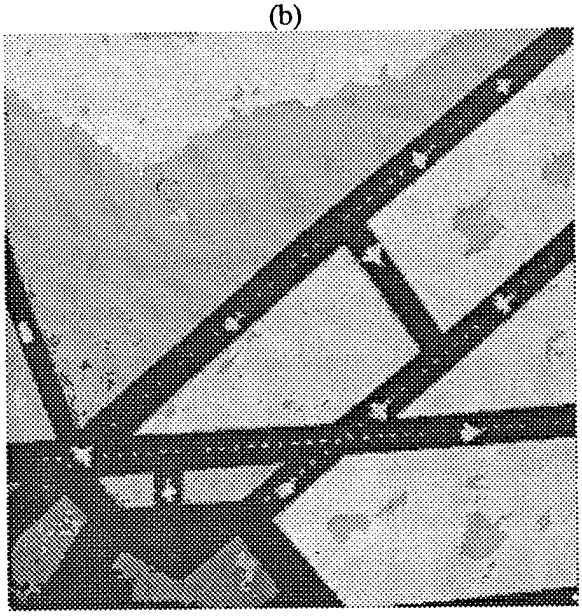
Fig. 2.6 Typical low resolution image and its corresponding segmentation images.



(a)



(b)



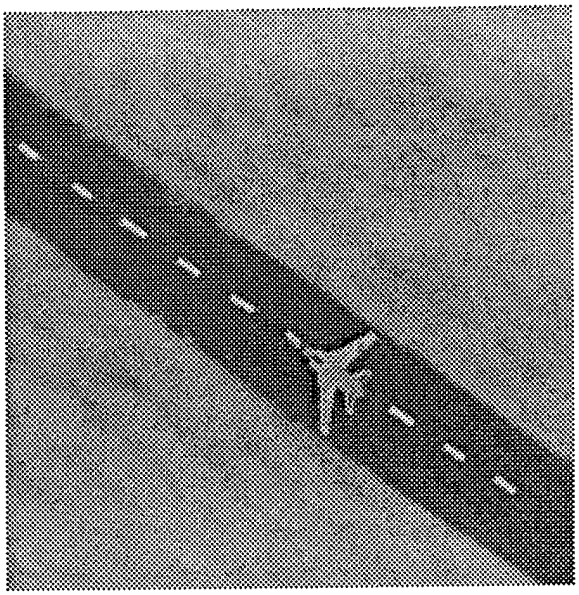
(c)

2.2.1 NMSU Scene Board Results

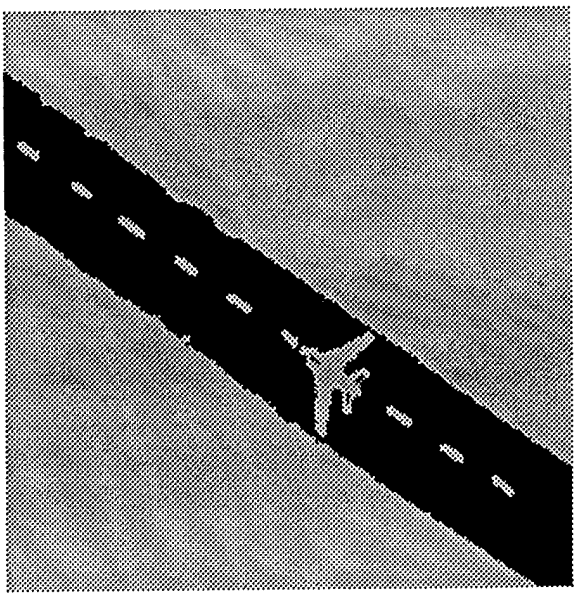
The NMSU scene board problem involves locating and identifying military targets in a desert airport landscape using our scene board. The scene board has a low-resolution monochromatic camera that views the whole scene board. A high resolution RGB camera views a small field that is controlled by a mirror control system. The low resolution image from the monochromatic camera is analyzed to locate areas likely to contain targets. The mirror control system positions the mirror so that the high resolution RGB camera can view selected regions. The high resolution color image is used to precisely locate and identify the objects of interest.

Scenes from our scene board are used to demonstrate the segmentation performance on low resolution monochromatic images and high resolution color images. The lighting conditions are uniform and the mixture distribution generally satisfy the Gaussian assumption. The maximum likelihood and gradient separation methods are used to locate objects of interest in low resolution airport scenes.

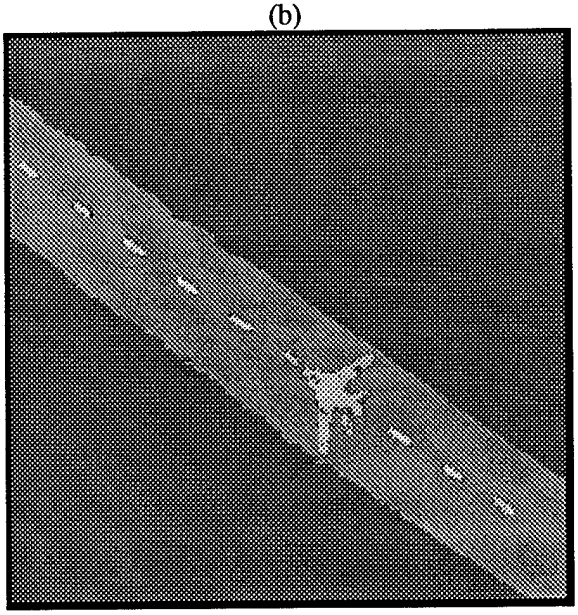
Fig. 2.7 High resolution image and its mixture components.



(a)



(b)

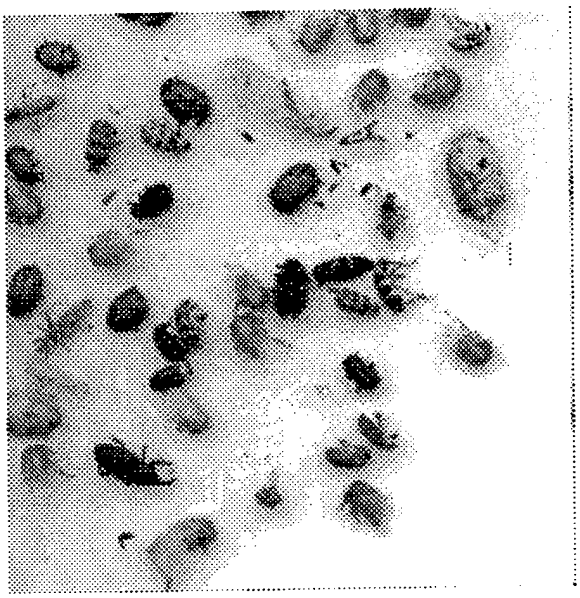


(c)

The major mixture components are separated, the scene components segmented and the objects are located as holes in the scene components. Fig. 2.6(a)(b)(c) shows a typical low resolution scene taken from the scene board and the segmentation results. The MLE and Bayesian decision rule were used to segment the scene into the scene components corresponding to the mixture components as shown Fig. 2.6(b). The gradient separation method was used to perform the mixture separation with similar results as shown in Fig. 2.6(c).

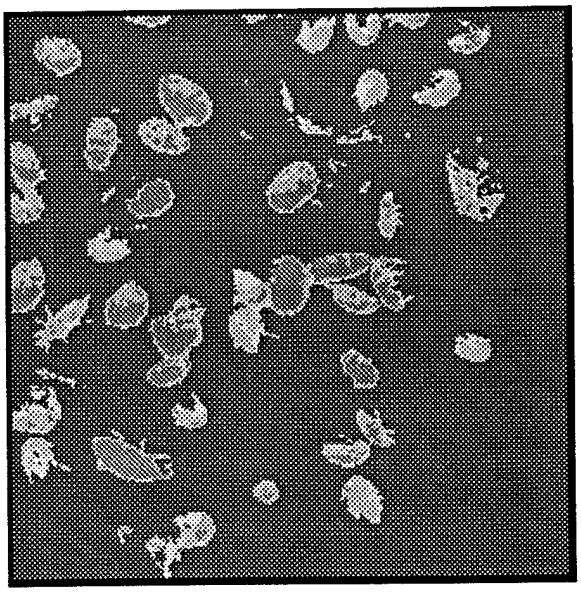
After the objects are located using the low resolution camera, the high resolution color camera was used to obtain a detailed description of the size, color and shape of each object. Three color images (red, blue, green) are generated and the segmentation methods used to segment the targets from the background as shown in Fig. 2.7(a)(b)(c). The red plane of the color image is shown in Fig. 2.7(a), the maximum likelihood segmentation is shown is Fig. 2.7(b) and the gradient segmentation is shown in Fig. 2.7(c).

Fig. 2.8 Insect image and its segmentation image



(a)

(b)



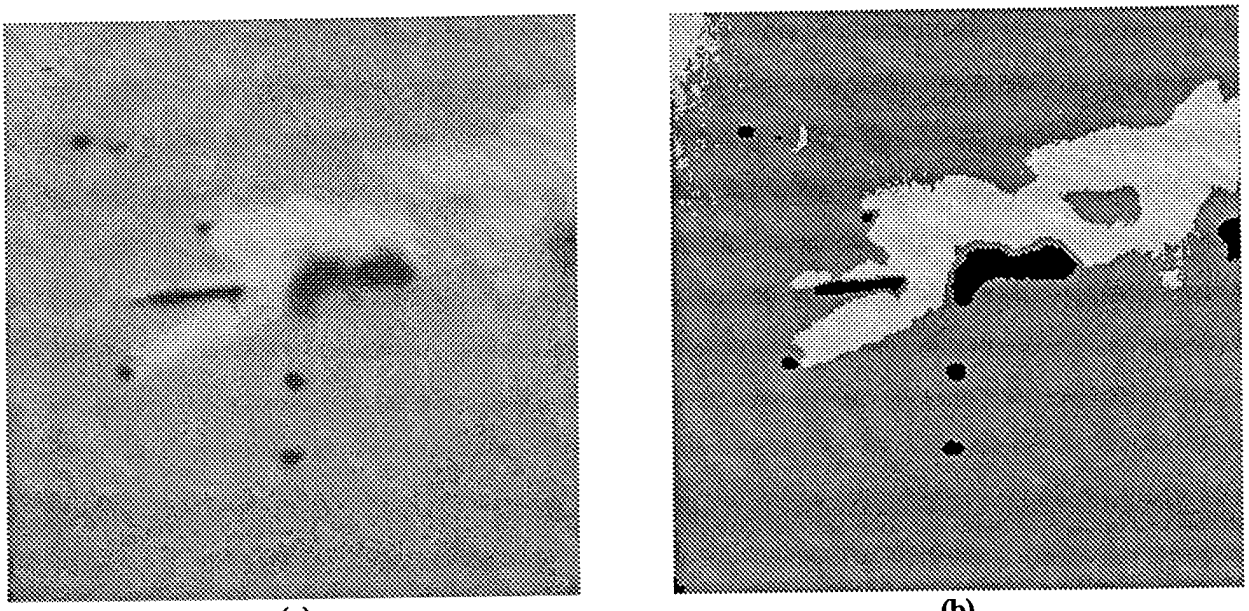
(c)

2.2.2 Insect Separation and Recognition Results

The insect problem demonstrates the application of the mixture separation methods in a multiple sensor environment. A high resolution color flatbed scanner was used to obtain red, blue and green insect images. The problem here is to separate and identify the insects using color and shape features. The lighting conditions are generally not uniform, shadows are present and the mixture components often do not satisfy the Gaussian assumption. The segmentation methods were used to separate the insect as shown in Fig. 2.8(a)(b)(c). The red plane of the color image is shown in Fig. 2.8(a), the maximum likelihood segmentation is shown is Fig. 2.8(b) and the gradient segmentation is shown in Fig. 2.8(c).

2.2.3 Missile and Submunition Separation Results

Finally, a flatbed scanner was also used to scan 70mm film sequences to obtain monochromatic missile and submunition images. The lighting conditions are not uniform and the mixture component distributions are Gaussian only over small regions. The segmentation methods were applied to separate the missile and submunitions from monochromatic images. The results are good as long as the missile and submunition are large and lighting conditions are uniform as shown in Fig. 2.9



(a) (b)

Fig. 2.9 (a) close view image; (b) corresponding mixture separated image

2.3 Multiple Object Tracker

The multiple object tracker resolves the nonuniform lighting problems by controlling the image resolution so that targets are small compared to scene components and using a flying spot scanner with a small window. By controlling the image resolution to keep the target size small, the color and texture distribution in a small window are unimodal and the visible targets are located on the tails of the distributions. The basic idea is to keep the target size large enough to adequately describe its shape but not its interior details.

The multiple object tracker consists of a window of interest controller (WC), a flying spot locator (FSL) and a multiple trajectory analyzer (MTA) as shown in Fig. 2.10. The window controller controls the image format resolution and defines the regions likely to contain objects of interest. Initially, the WC searches the global scene to locate possible targets. After objects are located and valid trajectories

established, the WC restricts the region of interest to predicted positions of targets. The flying spot locator locates the possible objects of interest on a frame-by-frame basis. The trajectory analyzer allocates the objects to trajectories and searches for the meaningful trajectories. Windows containing the predicted locations of the objects are returned to the window controller to limit the search on a frame by frame basis. Signals are generated by the trajectory analyzer to control the false alarm rate of the flying spot locator. Based on these signals, the FSL adjust the thresholds on the distribution tails to either increase or decrease the false alarm rate.

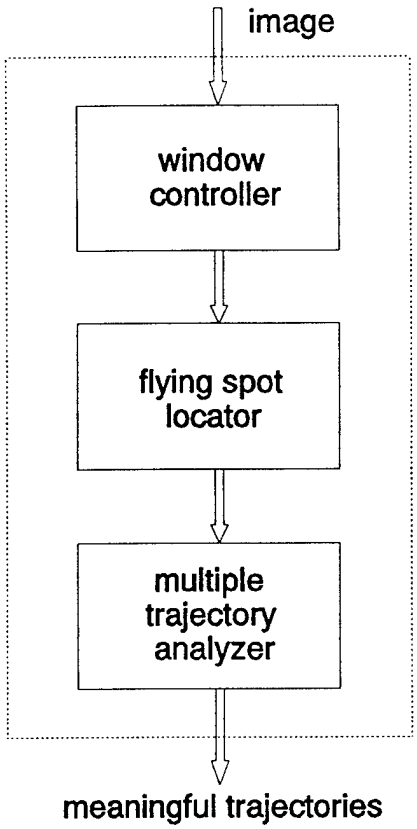


Fig. 2.10 Multiple object tracker.

2.3.1 Window Controller

The window controller is responsible for controlling the resolution and region of interest for the flying spot in an image. The resolution of the image format is controlled so that the objects of interest are small compared to the background scene components. The desired resolution is sufficient to describe the shape but not the interior details of the objects of interest. This requirement promotes the assumption that the color and texture distributions have one major mode corresponding to the local background and visible targets are located at the tail of the distributions. The second function of the window controller is to define the region of interest. Initially, the flying spot scans the whole image searching for targets of interest. When targets are found, the region of interest is restricted by the multiple object tracker to regions containing possible targets of interest. Thus, the flying spot does not scan an entire image and only scans the windows of interest in which the targets of interest may be located based upon the positions of targets in previous frame and their previous motion. This greatly increases the speed of the tracker.

2.3.2 Flying Spot Locator

The goal to develop a reliable and real-time oriented target tracking algorithm for complex scenes led to the development of the flying spot locator. The FSL consists of five modules: the flying spot scanner, the spot segmenter, the object locator, and the target recognizer as shown in Figure 2.11. The flying spot scanner provides a raster scan of the pixels in the window of interest generate by the window controller. The spot segmenter is based on the assumption that the distribution of the color and texture measurements is unimodal in a small region around the flying spot and the objects of interest are located on the tail of the distribution. The spot segmenter module estimates the mean and variance of the background features, and adaptively generates the tails of the distributions as the flying spot is scanned. The spot segmenter detects potential target pixels if the color or texture features are in the tails of the feature distributions. When a potential target is completely segmented from its background, the object locator estimates the size, color, and shape features. The target recognizer uses these features to identify the targets of interest. Thus, the targets of interest are found by the flying spot locator concurrently with the scanning process.

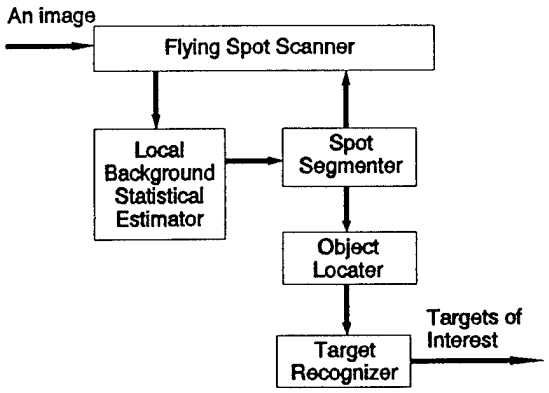


Fig. 2.11 The FSL structure.

The spot segmenter estimates the color and texture means and variances of the background pixels near the flying spot and generates the tail areas of the local color and texture distributions as the flying spot is scanned. These tail thresholds are adjusted to achieve a desired false alarm rate. This desired false alarm rate provides confidence that visible target pixels will be detected. When a possible target is detected the background estimator is turned off until the flying spot returns to the background scene.

The object locator is responsible for obtaining size, color, and shape of all objects detected. If a pixel is segmented as a potential target point by the spot segmenter, the object locator checks whether the target point is a new object or is 8-connected to an object currently being scanned. If a pixel combines two current objects then they are merged into one as shown in Fig. 2.12. At the end of each scan, the object locator checks to see if any objects are completely scanned and are ready for recognition analysis.

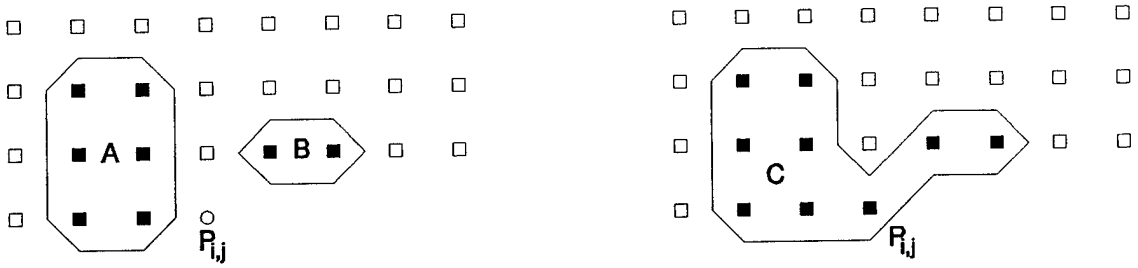


Fig. 2.12 Adjacent objects. (a) before $P_{i,j}$ processed by the FSL. (b) after $P_{i,j}$ processing.

After the target locator finds an object that is completely scanned and estimates its features, the target recognizer compares the features with that of the targets of interest. If they match, the object is considered as a target. Otherwise, the object is discarded.

Several images from different film sequences are used to illustrate the performance of the flying spot locator. The digitized film images were acquired using the Hewlett-Packard flatbed scanner with a transparency adapter. The optical resolution of the scanner is 400 pixels per inch. Figure 2.13 and 2.14 show the film image and their corresponding segmentation image after using FSL.

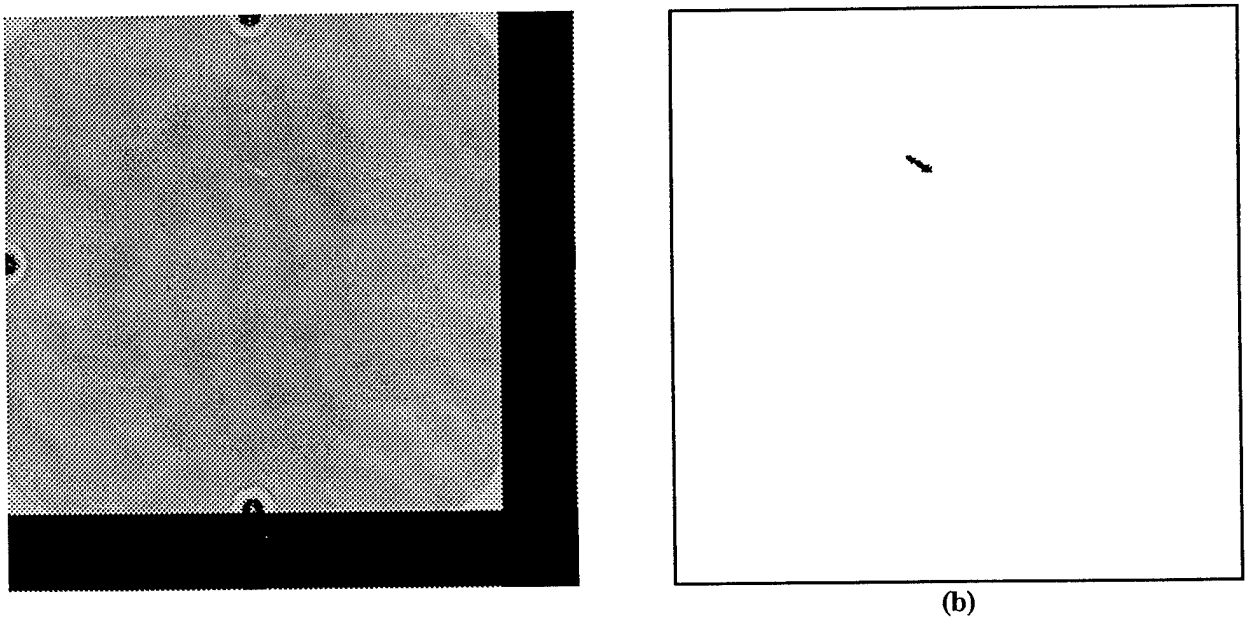
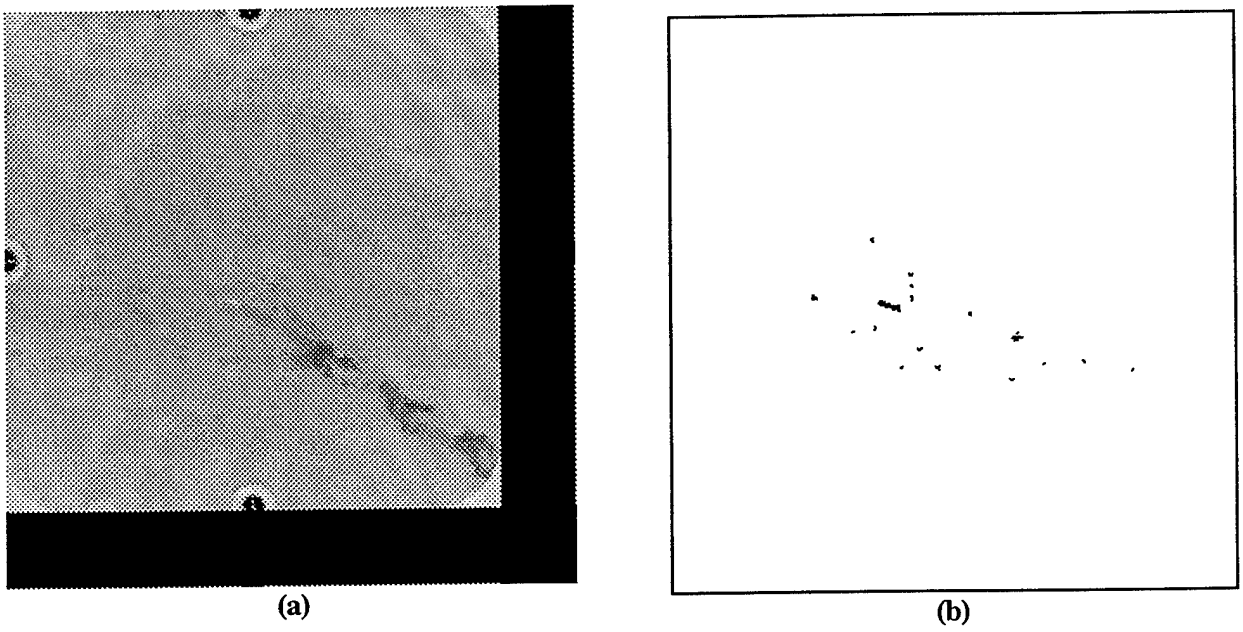


Fig. 2.13 (a) The image contained a carrier vehicle only; (b) the corresponding segmented image.

The modules of the flying spot locator all work on a pixel-by-pixel basis. Thus, objects are located and recognized when they are completely scanned. Since the modules all work in a pipeline-fashion, the FSL can be implemented with a parallel processing pipeline architecture.

Fig. 2.14 (a) Image containing carrier vehicle and submunitions; (b) segmented image.



(a) (b)

2.3.3 Multiple Trajectory Analyzer

An object's trajectory provides valuable information to assist in identifying objects with similar shape and color. The task of the multiple trajectory analyzer (MTA) is to analyze the size, color and shape of objects located on each frame, form possible trajectories and select the meaningful trajectories to be tracked.

The position of the targets is a function of the target motion, camera motion and other random factors involve in acquiring the images. Furthermore, with only one camera, objects are mapped into a plane and two or more targets may cross paths in the line of sight of the camera. These random factor cause the objects to appear to randomly jump and cross each other on a frame-to-frame basis, making the tracking problem very difficult even for the human visual system. The tracking problem is illustrated in Fig. 2.15 with a sequence of frames with targets. The basic problem is to locate the same target in the frame sequence given that the targets are randomly jumping and crossing each other and surrounded by other similarly shaped objects.

The MTA is implemented with a finite state machine structure as described in Section 2.1.4. On

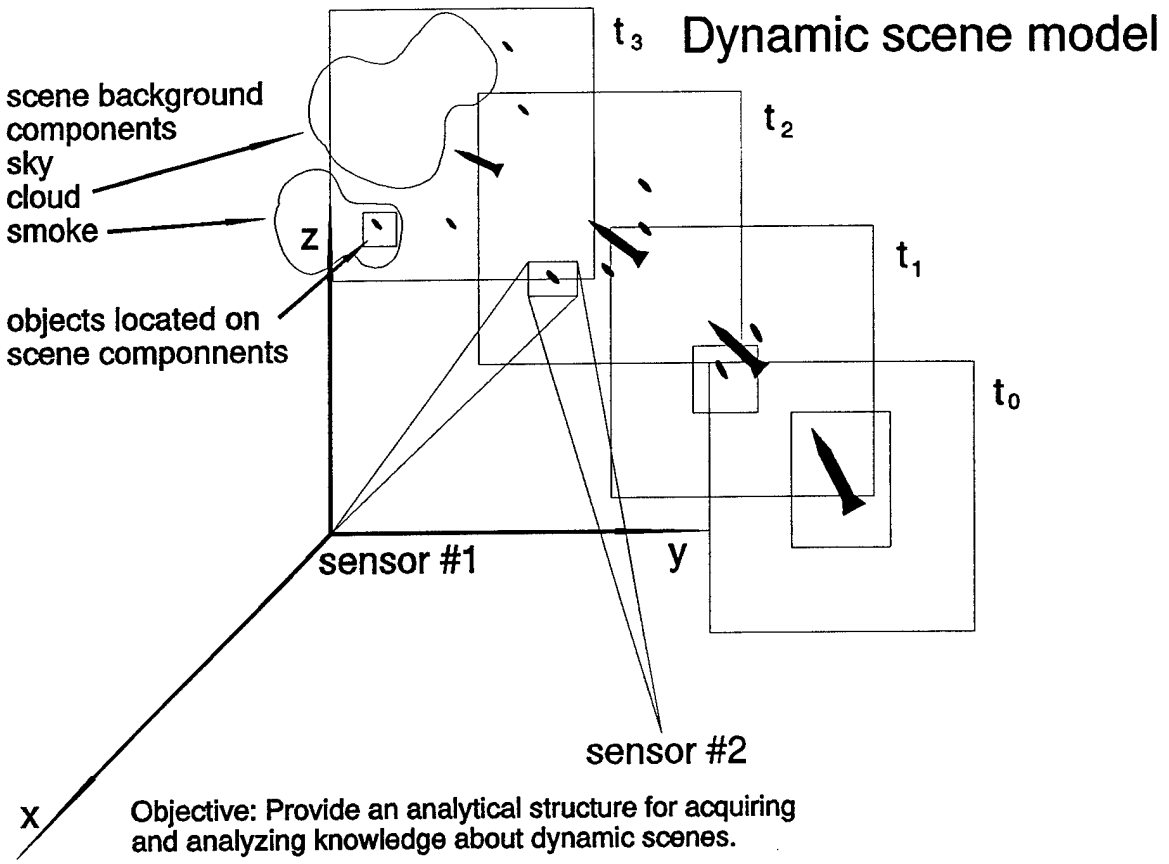


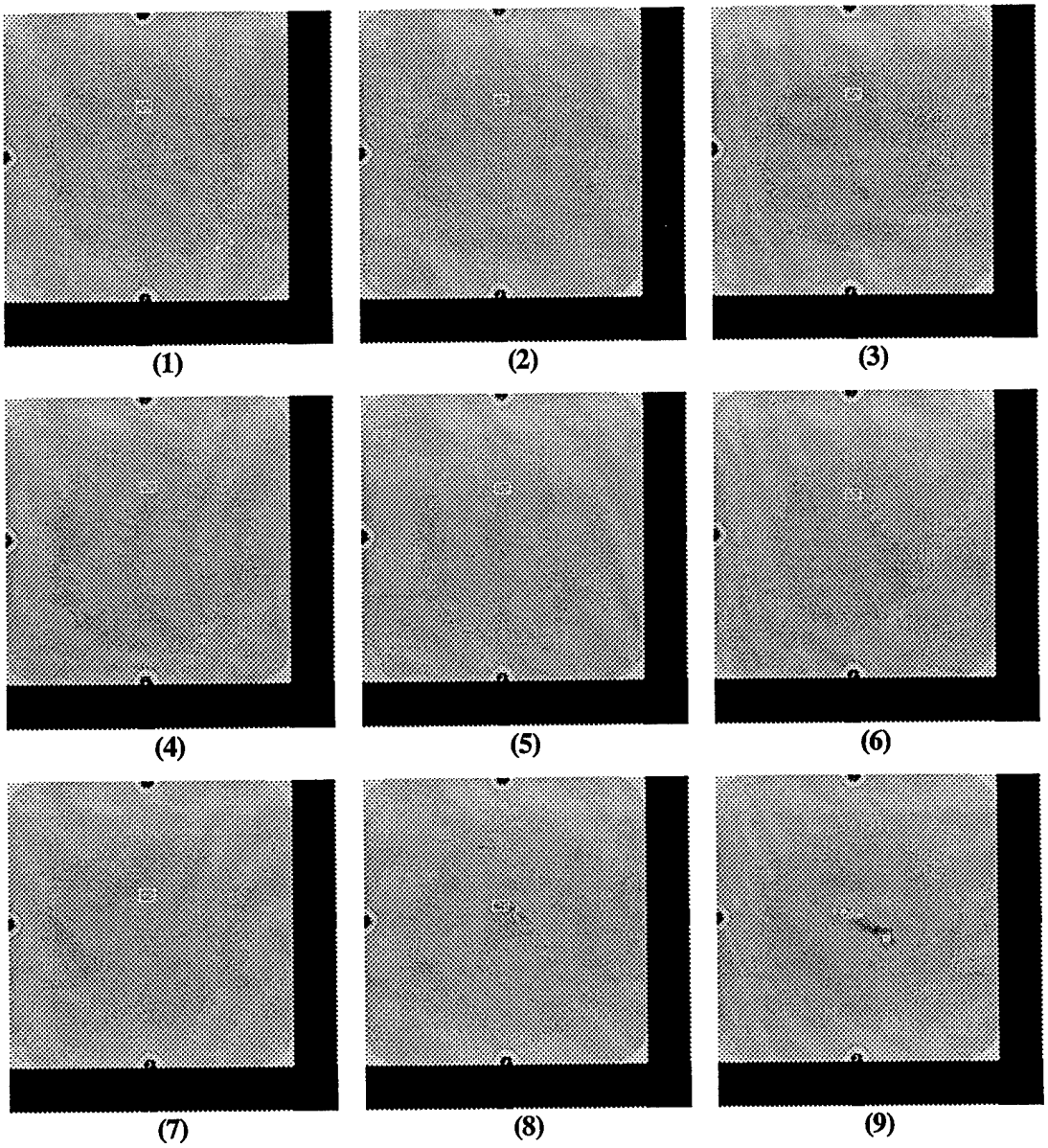
Fig. 2.16 Conceptualization of Dynamic Scene Model

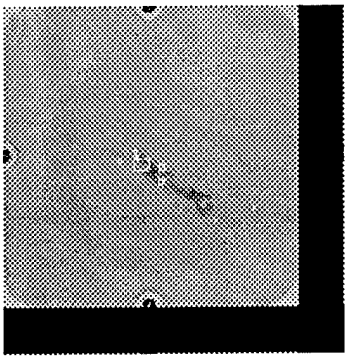
each frame the locator presents a list of possible targets and a description of the size, color and shape of each object. After two frames of data are obtain the MTA begins forming possible trajectories and predicting the next position of the targets. When the next list of possible targets arrives, the MTA compares the actual position with the predicted. Each time there is a good match between the actual and predicted position the confidence in the given trajectory is increased. When a trajectory predicts the position of a target five consecutive times it is elevated to a valid trajectory status by the state machine. If a trajectory fails to predict the location of the object, the trajectory is continued with the predicted position, the object is presumed to be occluded and confidence in trajectory is reduced. If the confidence in a trajectory falls below a given level the trajectory is terminated. During the process of tracking the trajectories, the MTA continues to search for new trajectories,using objects located but not claimed by a valid trajectory.

The MTA has been used to find the trajectories of the munitions fired from a missile in the 70mm film rolls. Since the munitions are very small, the resulting film images, which mark all possible targets

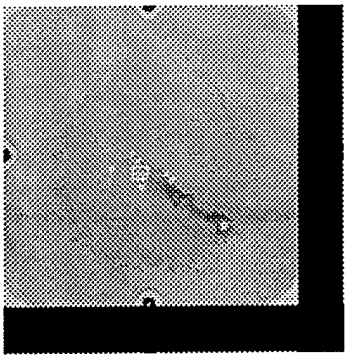
by a rectangle window with gray level 255, are shown in Fig. 2.16 (1)-(24). It can be seen from Fig. 2.16 that the munitions are not fired until the 9th frame. Since erratic motion is introduced by the tracking mount, the positions of munitions appear to vibrate. To remove erratic motion, the missile is used as a reference point. Since the locator can identify the missile from the munitions, the position of the missile is always defined as the origin of the coordinate system in the research. Thus, based upon the missile position, the relative munitions

Fig. 2.16 Sequence of 70mm images

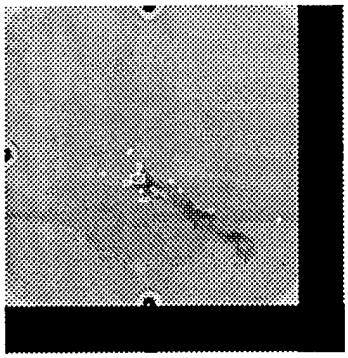




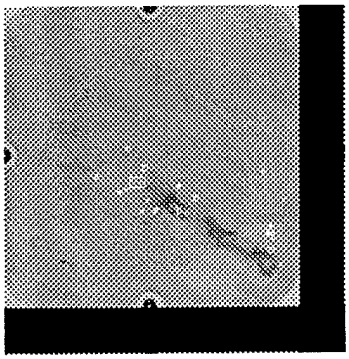
(10)



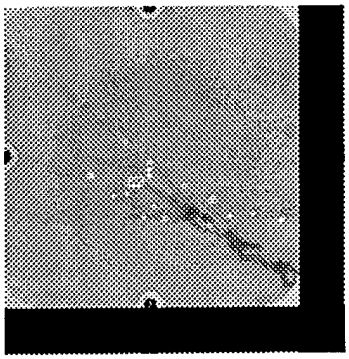
(11)



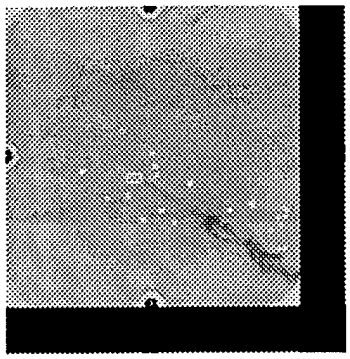
(12)



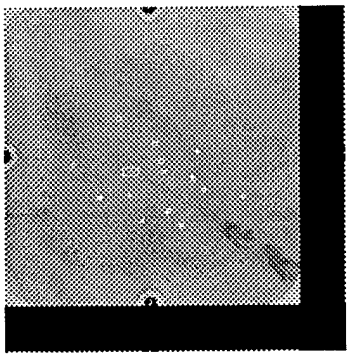
(13)



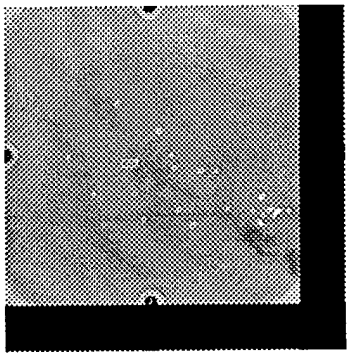
(14)



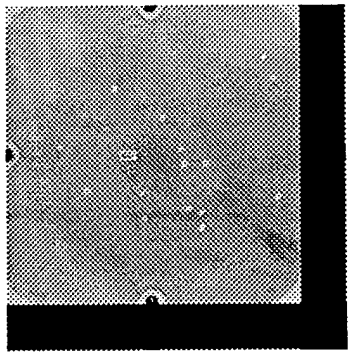
(15)



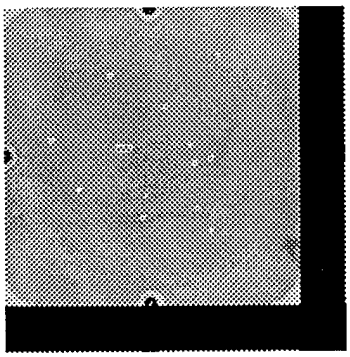
(16)



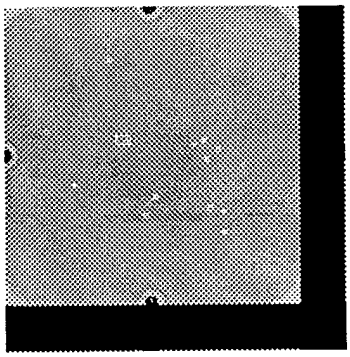
(17)



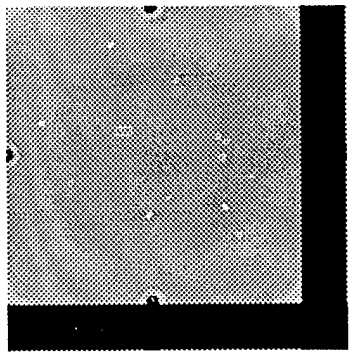
(18)



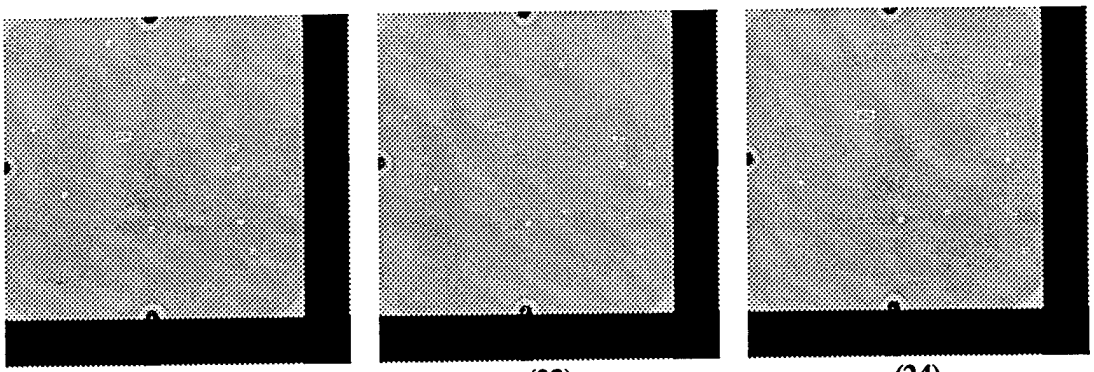
(19)



(20)



(21)



(22) (23) (24)

positions are determined. Fig. 2.17 shows all meaningful trajectories, where all target points are printed by dark squares and the predicted points are printed by gray squares. If a trajectory does not find its target in the current frame, a predicted location is used and these predicted locations are shown by dark dots. Although the locator provides many false objects, the trajectories in Fig. 2.17 are still very clear since the MTA only tracks meaningful trajectories. Fig. 2.18 shows the final meaningful trajectories by linking points of a trajectory in consecutive frames. In fact, all munitions trajectories in the frames of Fig. 2.16 (1)-(24) are found. A more difficult tracking sequence is given in Fig. 2.19. About 50% of the detections are false targets and many real targets are missed due to background clutter. The multiple trajectory analyzer still finds the meaningful trajectories among the observations as shown in Figs. 2.20 and 2.21. This demonstrates the remarkable ability of the MTA to handle really large false alarm and miss rates associated with low contrast targets and still find the meaningful trajectories.

Since all the calculations in the MTA are the fixed-point number calculations, the processing speed is very fast. When the MTA is used with a fast locator, the tracking process reaches 41 frames per second in a 120MHz Pentium PC.

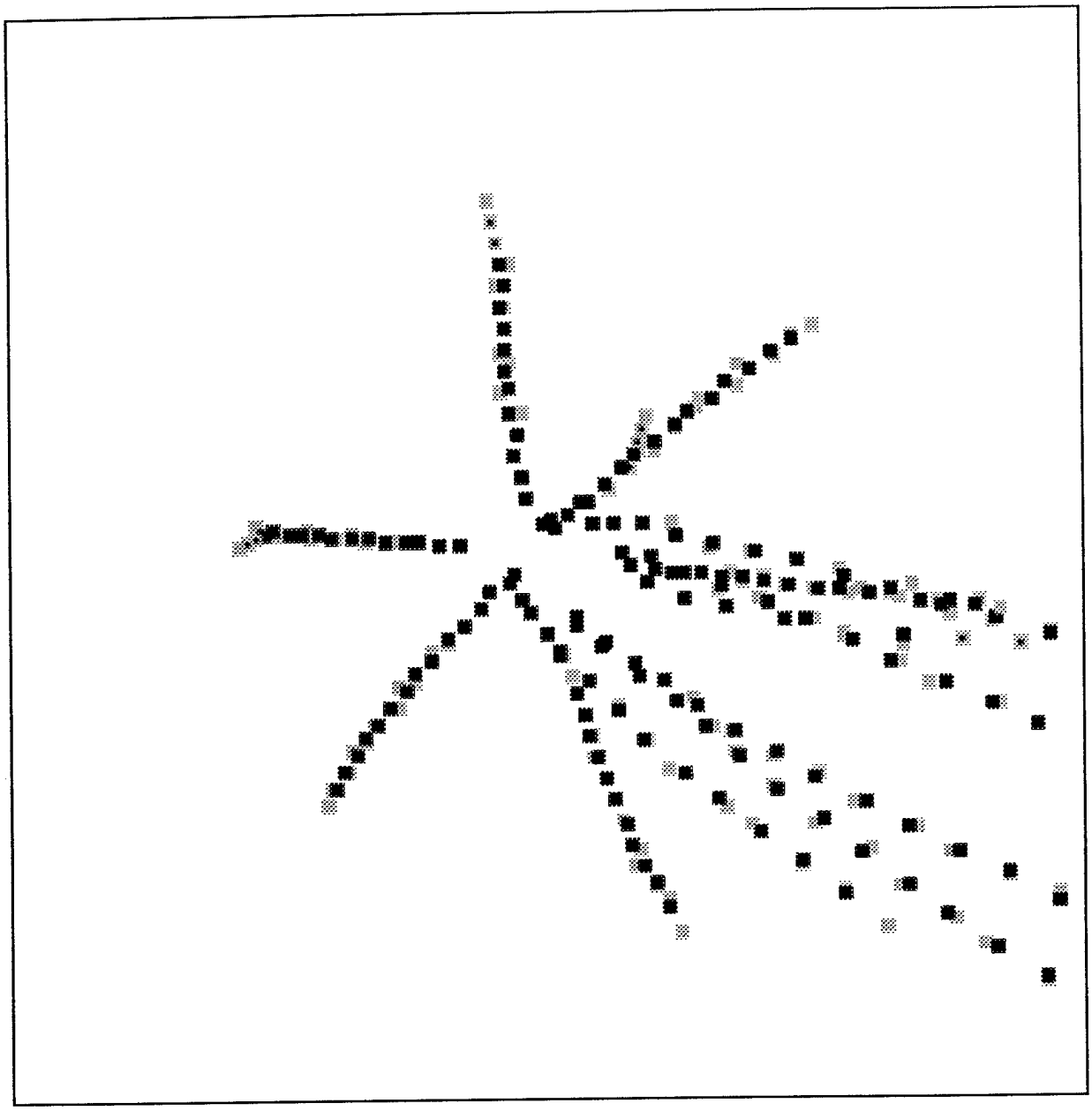


Fig. 2.17 All points in the meaningful trajectories.

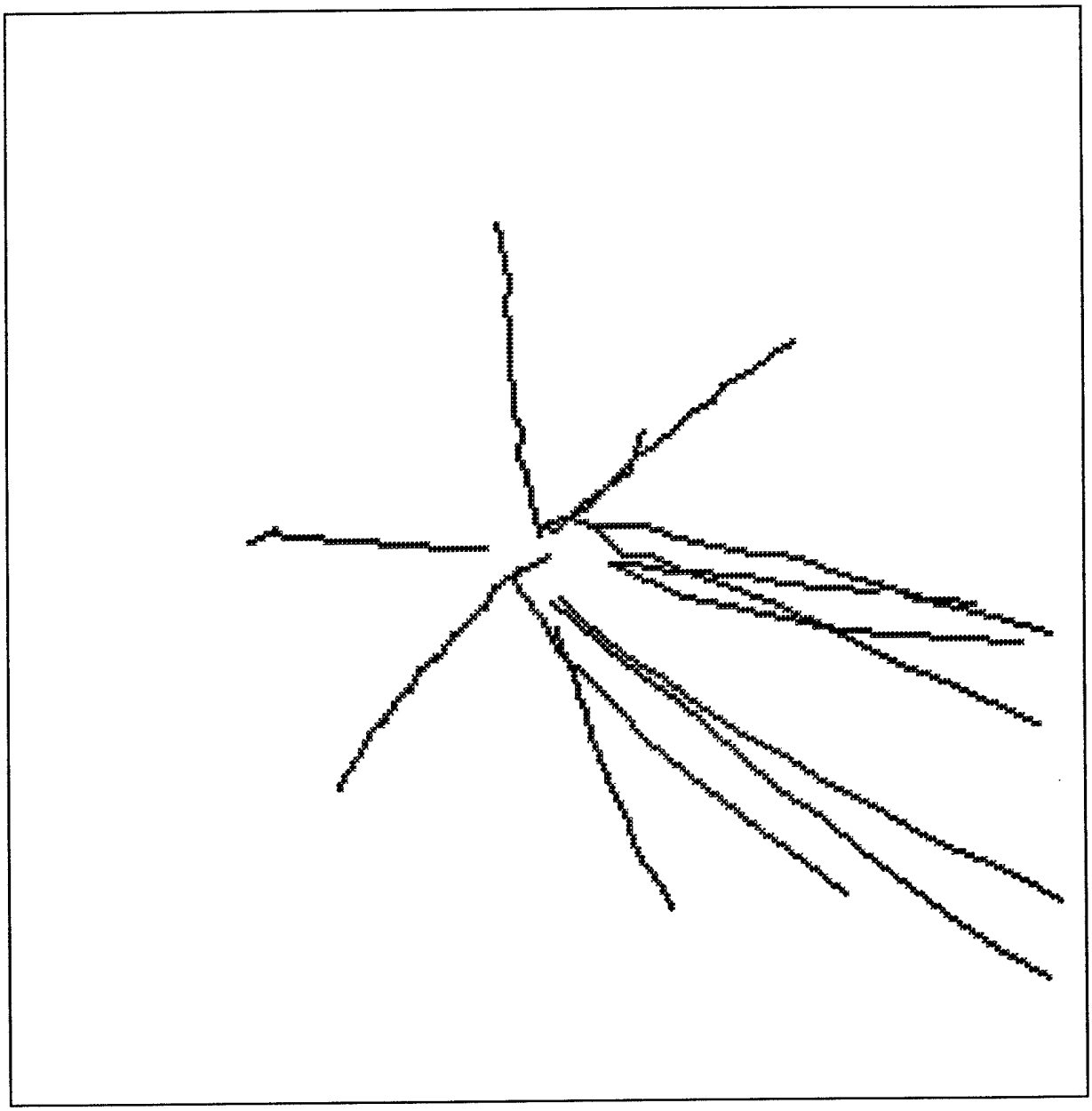
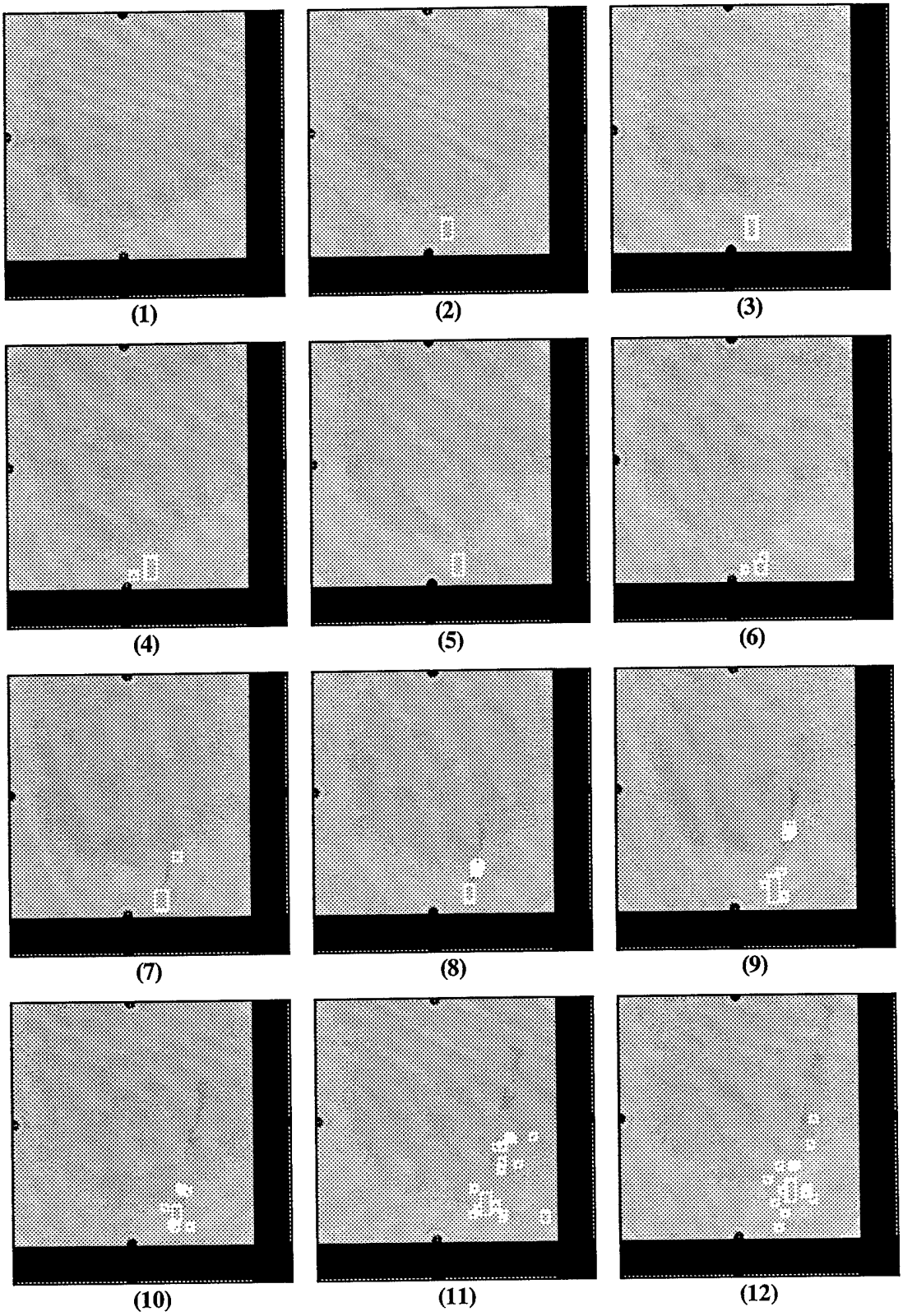
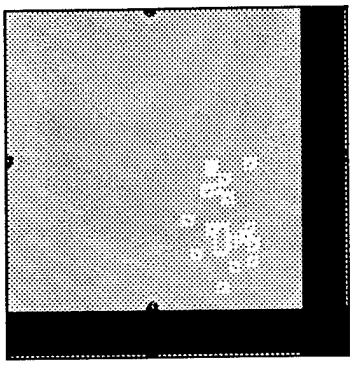


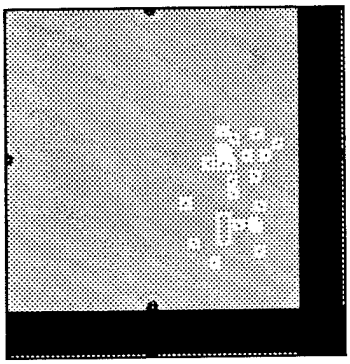
Fig. 2.18 All meaningful trajectories.

Fig. 2.19 Difficult sequence of 70mm images

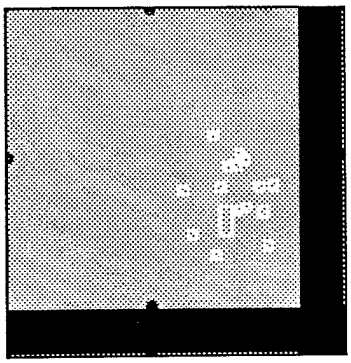




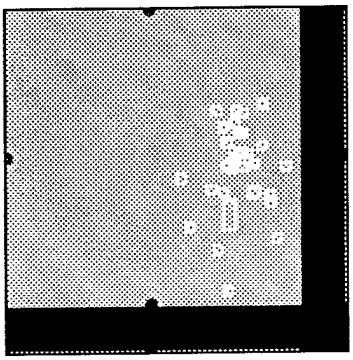
(13)



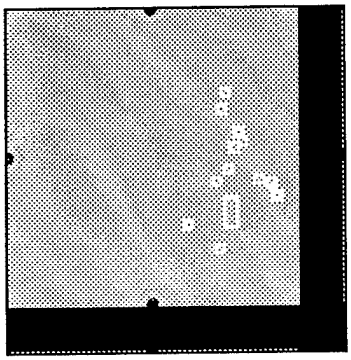
(14)



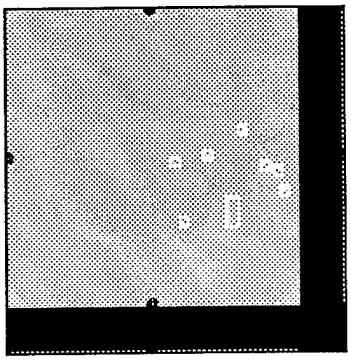
(15)



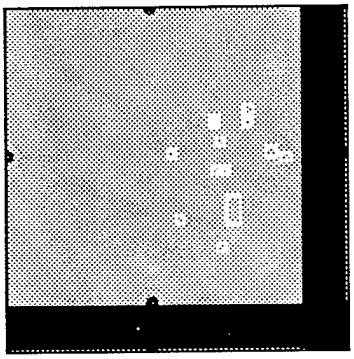
(16)



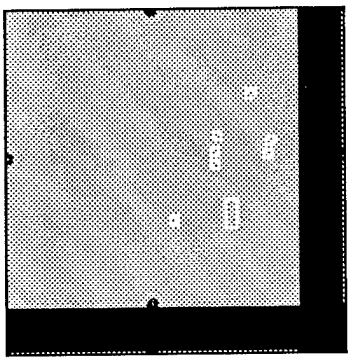
(17)



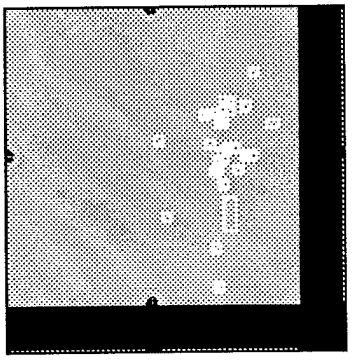
(18)



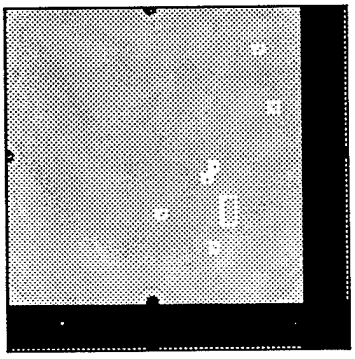
(19)



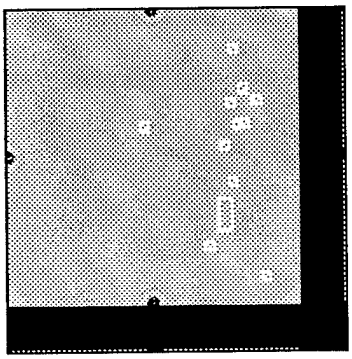
(20)



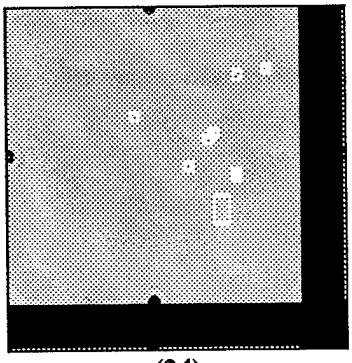
(21)



(22)



(23)



(24)

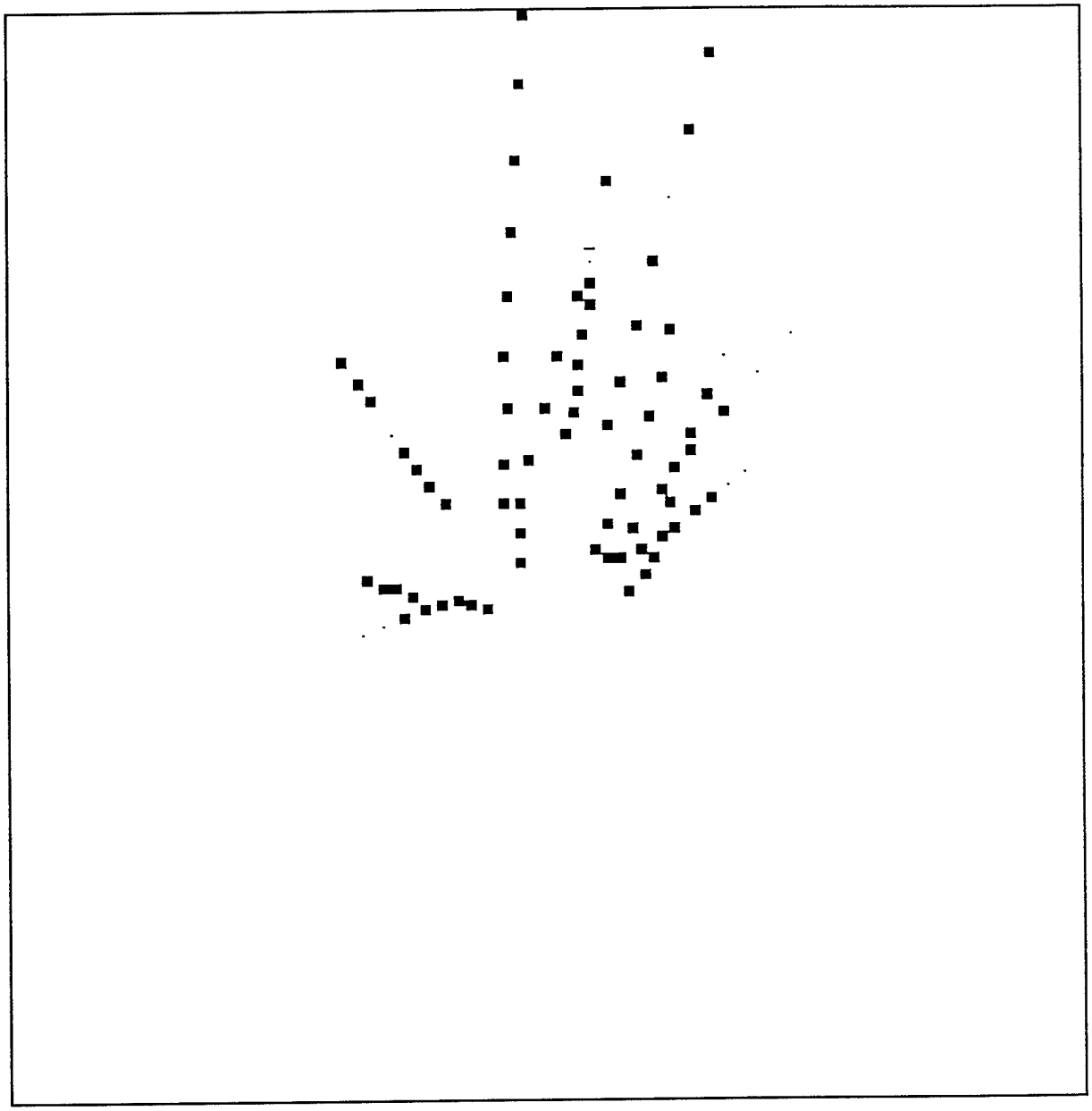


Fig. 2.20 All points in the meaningful trajectories.

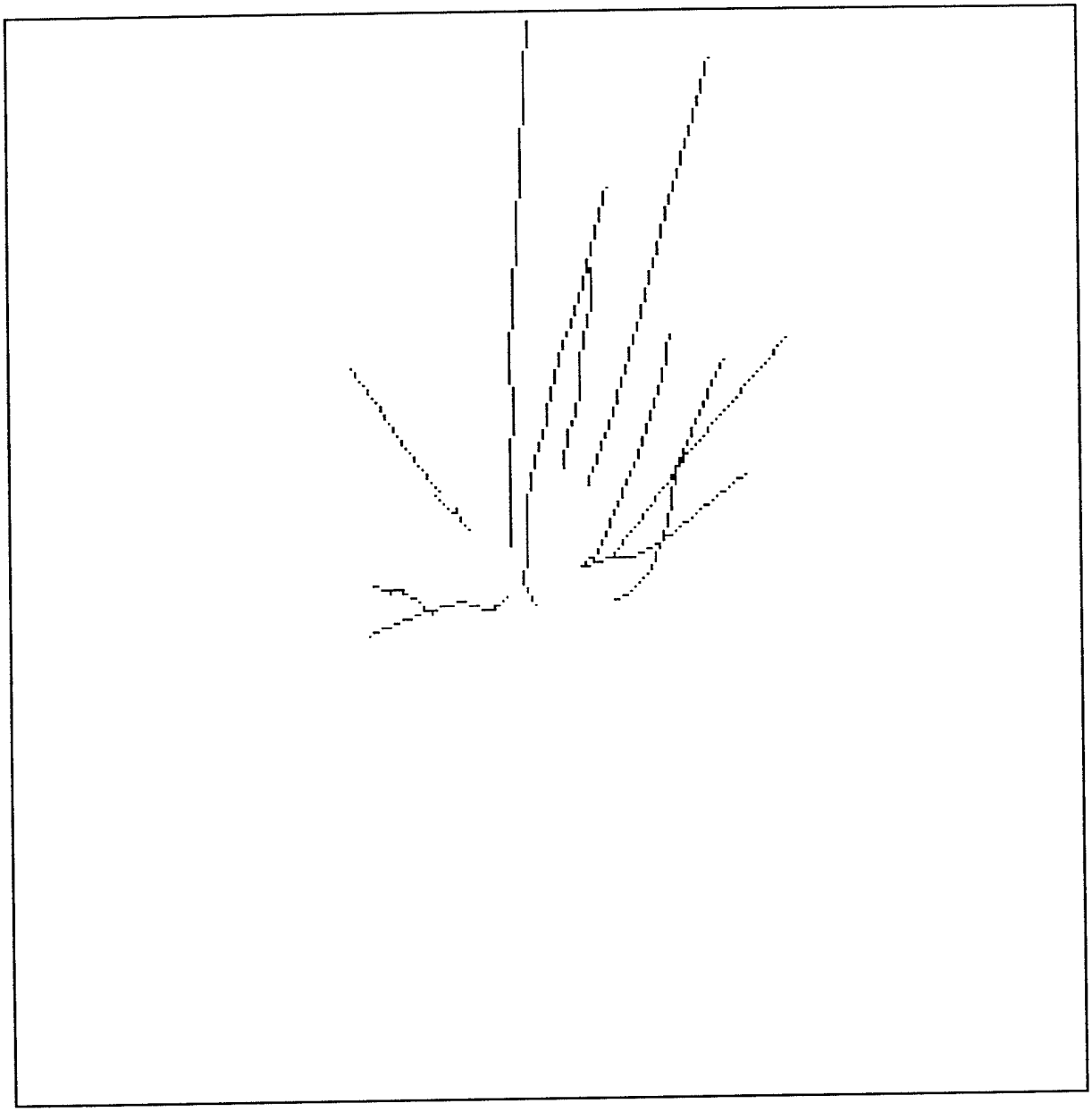


Fig. 2.21 All points in the meaningful trajectories.

2.3.4 Three dimensional Meaningful Tracking

The meaningful motion detector is used to locate objects that are of interest in each camera's field-of-view. The output of each camera is digitized into an image of MxN pixels. To establish three dimensional trajectories, the 3D position and pointing angles of the optical axis of each camera are required. Knowing the position and pointing angles of each camera, each pixel on an object generates a 3D pointing vector toward the object of interest. By intersecting these pointing vectors from multiple cameras, objects can be located, identified and tracked on a frame-to-frame basis.

The pointing vectors for each pixel in a camera's view are stored in two MxN matrices, $[\theta_{ij}]$ and $[\phi_{ij}]$, defining the azimuth (pan) and elevation (tilt) pointing angles for each pixel relative to the optical axis of the camera. Thus, a 3D object position can be determined from the pixel pointing vectors of two or more cameras.

Assume that the azimuth and elevation pointing angles of the optical axis of the camera are θ and ϕ, respectively. The azimuth and elevation pointing angles for a pointing vector passing through pixel (i,j) are $\theta+\theta_{i,j}$ and $\phi+\phi_{i,j}$, respectively. If the position of the camera (optical center) is (x, y, z) and the yaw angle is ω, then the pixel axis passing through pixel (i,j) is given by

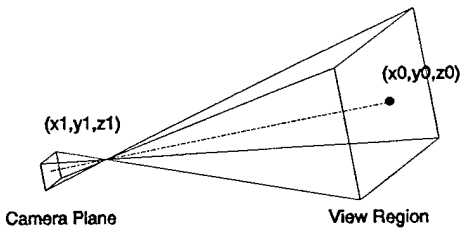


Fig. 2.22 The camera view.

$$\begin{cases} \dfrac{y_{i,j}-y}{x_{i,j}-x} = \tan(\theta+\arctan(\cos(\omega)\tan(\theta_{i,j}) \\ \qquad -\sin(\omega)\tan(\varphi_{i,j}))) \\ \dfrac{z_{i,j}-z}{\sqrt{(x_{i,j}-x)^2+(y_{i,j}-y)^2}} = \tan(\varphi+\arctan(\cos(\omega)\tan(\varphi_{i,j}) \\ \qquad +\sin(\omega)\tan(\theta_{i,j}))) \end{cases} \quad \textbf{(16)}$$

Therefore, the object found at pixel (i,j) of the image is at $(x_{i,j}, y_{i,j}, z_{i,j})$ in 3D space, which lies on or close to the pointing vector passing through (i,j) determined by equation (16). By adding a second camera and finding the intersection of the two pointing vectors from the two cameras, the 3D object position can be determined. The procedure is as follows.

Assume that (x_1, y_1, z_1) is the position of the master camera with optical axis azimuth angle θ_1

, elevation angle ϕ_1, and yaw angle ω_1, and (x_2, y_2, z_2) is the position of the slave camera with optical axis azimuth angle θ_2, elevation angle ϕ_2, and yaw angle ω_2, respectively. Furthermore, let (i_1, j_1) be an object position found by the master camera and (i_2, j_2) be an object position found by the slave camera, respectively. The pointing vector of pixel (i_1, j_1), L1, of the master camera satisfies

$$\begin{cases} \dfrac{y_{i_1,j_1}-y_1}{x_{i_1,j_1}-x_1} = \tan(\theta_1 + \arctan(\cos(\omega_1)\tan(\theta_{i_1,j_1}) \\ \qquad -\sin(\omega_1)\tan(\varphi_{i_1,j_1}))) \\ \dfrac{z_{i_1,j_1}-z_1}{\sqrt{(x_{i_1,j_1}-x_1)^2+(y_{i_1,j_1}-y_1)^2}} = \tan(\varphi_1 + \arctan(\cos(\omega_1)\tan(\varphi_{i_1,j_1}) \\ \qquad +\sin(\omega_1)\tan(\theta_{i_1,j_1}))) \quad , \end{cases} \tag{17}$$

and the pointing vector of pixel (i_2, j_2), L2, of the slave camera satisfies

$$\begin{cases} \dfrac{y_{i_2,j_2}-y_2}{x_{i_2,j_2}-x_2} = \tan(\theta_2 + \arctan(\cos(\omega_2)\tan(\theta_{i_2,j_2}) \\ \qquad -\sin(\omega_2)\tan(\varphi_{i_2,j_2}))) \\ \dfrac{z_{i_2,j_2}-z_2}{\sqrt{(x_{i_2,j_2}-x_2)^2+(y_{i_2,j_2}-y_2)^2}} = \tan(\varphi_2 + \arctan(\cos(\omega_2)\tan(\varphi_{i_2,j_2}) \\ \qquad +\sin(\omega_2)\tan(\theta_{i_2,j_2}))) \quad , \end{cases} \tag{18}$$

By defining

$$\begin{cases} \theta^*_{i_1,j_1} = \arctan(\cos(\omega_1)\tan(\theta_{i_1,j_1}) - \sin(\omega_1)\tan(\varphi_{i_1,j_1})) \\ \varphi^*_{i_1,j_1} = \arctan(\cos(\omega_1)\tan(\varphi_{i_1,j_1}) + \sin(\omega_1)\tan(\theta_{i_1,j_1})) \quad , \end{cases} \tag{19}$$

equation (17) can be rewritten as

$$\begin{bmatrix} x_{i_1,j_1} \\ y_{i_1,j_1} \\ z_{i_1,j_1} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + t_1 \begin{bmatrix} \cos(\theta_1 + \theta^*_{i_1,j_1})\cos(\varphi_1 + \varphi^*_{i_1,j_1}) \\ \sin(\theta_1 + \theta^*_{i_1,j_1})\cos(\varphi_1 + \varphi^*_{i_1,j_1}) \\ \sin(\varphi_1 + \varphi^*_{i_1,j_1}) \end{bmatrix} \quad , \tag{20}$$

which is denoted as

$$r_{i_1,j_1} = r_1 + t_1 v_1 \ . \tag{21}$$

Similarly, by defining

$$\begin{cases} \theta^*_{i_2,j_2} = \arctan(\cos(\omega_2)\tan(\theta_{i_2,j_2}) - \sin(\omega_2)\tan(\varphi_{i_2,j_2})) \\ \varphi^*_{i_2,j_2} = \arctan(\cos(\omega_2)\tan(\varphi_{i_2,j_2}) + \sin(\omega_2)\tan(\theta_{i_2,j_2})) \end{cases}, \tag{22}$$

equation (18) can be rewritten as

$$\begin{bmatrix} x_{i_2,j_2} \\ y_{i_2,j_2} \\ z_{i_2,j_2} \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} + t_2 \begin{bmatrix} \cos(\theta_2 + \theta^*_{i_2,j_2})\cos(\varphi_2 + \varphi^*_{i_2,j_2}) \\ \sin(\theta_2 + \theta^*_{i_2,j_2})\cos(\varphi_2 + \varphi^*_{i_2,j_2}) \\ \sin(\varphi_2 + \varphi^*_{i_2,j_2}) \end{bmatrix}, \tag{23}$$

which is denoted as

$$r_{i_2,j_2} = r_2 + t_2 v_2 \ . \tag{24}$$

When the pointing vectors passing through (i_1 , j_1) of the master camera and passing through (i_2 , j_2) of the slave camera hit the same object in 3-D space, the distance between of L1 and L2 should be very small. By equations (21) and (22), the distance between L1 and L2 is

$$d = \frac{|(r_2 - r_1)\cdot(v_1 \times v_2)|}{|v_1 \times v_2|} \ . \tag{25}$$

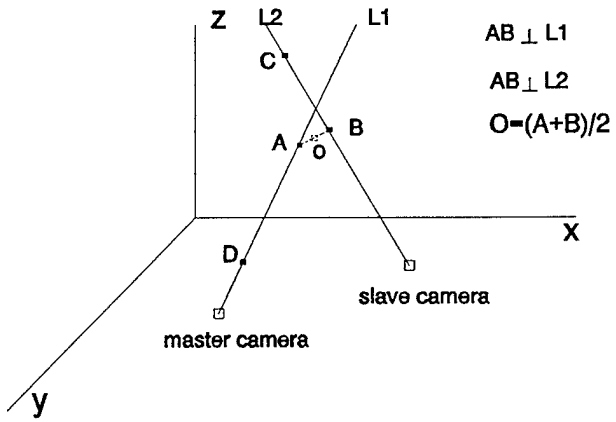


Fig. 2.23 Two pixel axes.

If L1 and L2 are close to one another (the distance of AB is small as shown in Fig. 2.23), the objects found by both the master camera and the slave camera could be the same object at position O in the 3D space, where

$$O = \frac{A+B}{2} \tag{26}$$

In fact, The position of O point can be obtained from equations (21) and (24) as:

$$O = \frac{r_1 + r_2 + t_{o1}v_1 + t_{o2}v_2}{2} \ , \tag{27}$$

where

$$t_{o1} = \frac{(v_1 \times v_2) \cdot [(r_2 - r_1) \times v_2]}{(v_1 \times v_2)^2} \quad , \tag{28}$$

$$t_{o2} = \frac{(v_1 \times v_2) \cdot [(r_2 - r_1) \times v_1]}{(v_1 \times v_2)^2} \quad . \tag{29}$$

Thus, an 3D object position is obtained from two pointing vectors. If more than two cameras are used the object position is obtained by finding a point that minimizes the sum of the squared distances to all pointing vectors.

The 3D tracker consists of five modules: (1) The 3D pointer module which locates the object using two or more pointing vectors, (2) a find module which continually attempts to find new trajectories, (3) a track module which selects objects to continue existing trajectories, (4) a merge module for merging trajectories, and (5) a finite state controller that selects the meaningful trajectories as shown in Fig. 2.24. The pointer module generates a 3D position from pointing vectors from two or more cameras. The merge module, the find module, and the finite state controller in the 3D tracker are the same as those in the 2D tracker.

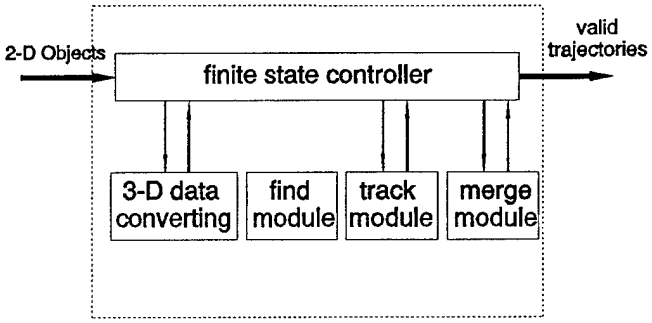


Fig. 2.24 The 3-D tracker.

The track module is responsible for finding a target for each existing trajectory. It consists of two parts. The first part, same as the track module in the 2D tracker, allocates objects to existing trajectories if they are the closest objects in the region of acceptance and have the proper shape and color. The second part attempts to locate objects for existing trajectories when only one camera sees the target. If a trajectory can not find any 3D object in the region of acceptance, the tracker module searches for a pointing vector from just one camera passing through the region of acceptance around predicted position A, as shown in Fig. 2.25. If a pointing vector L passes through the region of acceptance then the point B on the pointing vector closest to A is assigned to the trajectory.

By combining the meaningful trajectory idea with multiple cameras, the resulting 3D tracker is

capable of reliably tracking multiple objects in very complex environments. The system is currently being used to monitor activity in a room, using four cameras located near the ceiling in the four corners of the room. When an object is detected by two or more cameras, the monitoring system locates, tracks and provides a 3D description of the object of interest.

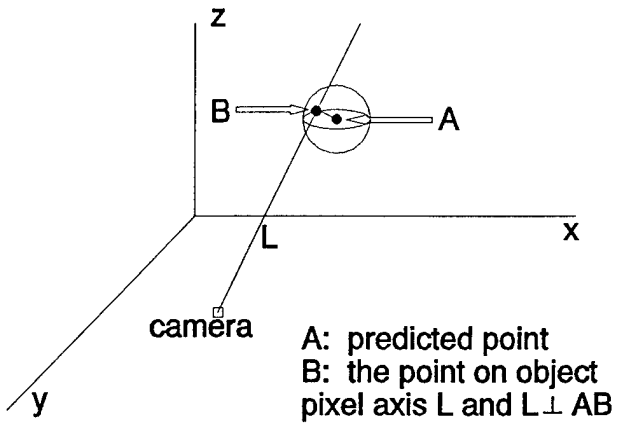


Fig. 2.25 One pixel axis predict.

2.4 References

[1] Papoulis, A., *Probability, Random variables and Stachastic Processes*, McGraw-Hill, New York, 1965.

[2] Jordan J. B. and Flachs G. M., "Statistical Segmentation of Digital Images," *Proceedings of SPIE*, Vol. 754, Optical and Digital Pattern Recognition, pp. 220-228, January 1987.

[3] Bendat, J., Piersol, A., *Measurement and Analysis of Random Data*, Wiley, 1966.

[4] Dude R., Hart, P., *Pattern Classification and Scene Analysis*, Wiley, 1974.

[5] Van Trees, H., *Detection, Estimation and Modulation Theory*, Wiley, 1971.

[6] Bao, Z., and Flachs, G.M., "A New Approach for Locating Multiple Targets," *The Proceeding of SPIE in Signal Processing, Sensor Fusion, and Target Recognition IV*, 1994.

[7] Bao, Z., Flachs, G. M., and Jordan, J. B., "Locating Multiple Targets In Complex Scenes," *The Proceeding of SPIE in Signal Processing, Sensor Fusion, and Target Recognition IV*, Vol.1684, 1995.

[8] Bao, Z., Flachs, G. M., and Jordan, J. B., "New Method for Tracking Multiple Trajectories," *The Proceeding of SPIE in Signal Processing, Sensor Fusion, and Target Recognition IV*, Vol.1684, 1995.

[9] Carlson, Jeffrey J., "Decision-Making Complexity with Applications in Electronic Vision," Ph.D. dissertation from New Mexico State University, Sept. 1988.

[10] Beer, Cynthia, "The Tie Statistic and Texture Recognition," Ph.D. dissertation from New Mexico State University, Dec. 1989.

[11] Flachs, G. M., Cynthia L. Beer and David R. Scott, "A Well-Ordered Feature Space Mapping for Sensor Fusion," *Proceedings of the SPIE 2nd International Conference on*

Sensor Fusion, vol. 1100, March 1989.

[12] Beer, C. L., G. M. Flachs, D.R. Scott, and J.B. Jordan, "Feature Selection and Decision Space Mapping for Sensor Fusion," *Proceedings of the SPIE 1990 Symposium on Sensor Fusion II*, Vol. 1198-19, Nov. 1989.

[13] Flachs, G. M., J. B. Jordan, C. L. Beer, and D. R. Scott, "Feature Space Mapping for Sensor Fusion," *Journal Robotic Systems*, Vol. 7, No. 3, pp. 373-393, June 1990.

[14] Scott, D. R., "The K-Nearest Neighbor Statistic With Applications to Electronic Vision Systems," Ph.D. dissertation from New Mexico State University, Dec. 1990.

[15] Scott, D. R., G. M. Flachs and P. T. Gaughan, "Sensor Fusion Using K-nearest Neighbor Concepts," *Proceedings of the SPIE Symposium on Advances in Intelligent Systems, Sensor Fusion III*, Boston, November 1990.

[16] Choe, H., "A comparative analysis of statistical, fuzzy, and artificial neural pattern recognition techniques," Ph.D. dissertation from New Mexico State University, 1992.

[17] Fukunaga, Keinosuke, *Statistical Pattern Classification*, Handbook of Pattern Recognition and Image Processing, Tzay Y. Young and King-Sun Fu (EDS), Academic Press, New York, 1986.

[18] Flachs, G. M. and Z. Bao, "Nearest neighbor controller," *Proceedings of the SPIE-1992 Conference on Signal Processing and Control*, Orlando Florida, April 1992.

[19] G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ, Princeton Univ. Press, 1976.

[20] Conover, W. J., *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., 1980.

[21] Titterington, D. M., Smith, A. F., and Makov, U. E., *Statisticala Analysis of Finite Mixture Distributions*, John Wiley & Sons, 1985.

[22] Everitt, B. S. and Hand, D. J., *Finite Mixture Distributions*, Chapman and Hall, 1981.

[23] Tou, J. T., and Gonzalez, R. C., *Pattern Recognition Principles*, Addison-Wesley Publishing Company, 1974

[24] Fugunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, 1990.

3. PUBLICATIONS AND REPORTS

[1] Bao, Z., "Locating and finding multiple meaningful trajectories," Ph.D. Dissertation from New Mexico State University, July 1995.

[2] Lewis, M., "Edge Detection at Subpixel Accuracy Using Truncated Interpolation," Ph.D. dissertation from New Mexico State University, August 1995.

[3] Khurrum, M., "Edge Representation Using Bezier Polynomials," Ph.D. dissertation from New Mexico State University, August 1995.

[4] Choe, H "A comparative analysis of statistical, fuzzy, and artificial neural pattern recognition techniques," Ph.D. dissertation from New Mexico State University, 1992.

[5] Wang, W."Hand recognition by Wavelet Packet Transform and Neural Network," MS Technical Report, 1995.

[6] Narsipur,D. "Digital Image Edge Compression Using Bezier Polynomials," MS Thesis, December 1994.

[7] Bao, Z., and Flachs, G.M., "A New Approach for Locating Multiple Targets," *The Proceeding of SPIE in Signal Processing, Sensor Fusion, and Target Recognition IV*, 1994.

[8] Bao, Z., Flachs, G. M., and Jordan, J. B., "Locating Multiple Targets In Complex Scenes," *The Proceeding of SPIE in Signal Processing, Sensor Fusion, and Target Recognition IV*, Vol.1684, 1995.

[9] Bao, Z., Flachs, G. M., and Jordan, J. B., "New Method for Tracking Multiple Trajectories," *The Proceeding of SPIE in Signal Processing, Sensor Fusion, and Target Recognition IV*, Vol.1684, 1995.

[10] Wang, W., Bao, Z., Meng, Q., Flachs, G. M., Jordan, J. B., and Carlson, J., "Hand recognition by Wavelet Packet Transform and Neural Network," *The Proceeding of SPIE in Signal Processing, Sensor Fusion, and Target Recognition IV*, Vol.1684, 1995.

[11] Lewis, M. and Jordan J. B., "Edge-data Compression Using Bézier Polynomials," *Proceedings of the SPIE-1994 Conference on Signal Processing, Sensor Fusion and Target Recognition II*, Orlando, Florida, April 1994.

[12] Jordan, J. B. and Watkins, W., "Determination of continuous system transfer functions from sampled pulse response data," *Optical Engineering*, Vol. 33, No. 12, Dec., 1994.

[13] Carlson, J., Flachs, G. M. and Jordan J. B., "Real-time 3D visualization of volumetric video motion sensor data,"xxxx.

4. PERSONNEL SUPPORTED AND DEGREES GRANTED

Major Investigators	Research Area
Flachs, G. M.	Image Processing
Jordan, J. B.	Image Processing

Doctoral Students	Dissertation
Bao, Z.	"Locating and finding multiple meaningful trajectories," Ph.D. Granted.
Lewis, M.	"Edge Detection at Subpixel Accuracy Using Truncated Interpolation," Ph.D. Granted.
Khurrum, M.	"Edge Representation Using Bezier Polynomials," Ph.D. Granted.
Choe, Howan	"A comparative analysis of statistical, fuzzy, and artificial neural pattern recognition techniques," Ph.D. Granted

M.Sc. Students

Wang, W.	"Hand recognition by Wavelet Packet Transform and Neural Network," M.S.E.E. Granted
Narsipur,D.	"Digital Image Compression Using Bezier Polynomials," M.S.E.E. Granted

Undergradate Students Supported

Hardin. C.	B.S. Electrical and Computer Engineering, Anticipated May, 1997

5. GOVERNMENT/INDUSTRIAL CONTACTS

An important goal of the research project is to transfer new concepts and techniques to the Army laboratories and industry. Often this requires the investigators to assist the laboratories implement the concepts in their applications. The feedback, however, is important in evaluating and motivating the basic research. Some important technology transfer activities related to the sponsored research are:

* The image processing research group assisted the Atmospheric Science Laboratory's Target Contrast Characterizer project team in establishing the continuous system transfer function for their IR sensors.

* The image processing research group assisted the Instrumentation Division at White Sands Missile Range design a near real-time film reading processing system using the results of the research sponsored by this research contract.

* The image processing research group together with the NMSU entomology department developed an color vision system for automatic recognition of cotton insects.

* The image processing research group assisted Sandia National Laboratory design a biometric human verification system for security entry control.

* The image processing research group assisted Sandia National Laboratory design a 3D volumetric video motion system for monitoring the movement of personnel and materials.

Technical contacts made during the grant are:

NAME	ORGANIZATION	APPLICATION
Jeffrey Carlson	Sandia National Labs	3D Tracking Applications
Steven Ortiz	Sandia National Labs	Human Verification System
Daniel Pritchard	Sandia National Labs	Accoutics Surveillance System
Wendell Watkins	Atmospheric Sciences Lab.	IR Image Complexity
Joe Ambrose	WSMR Instrumentation Div.	Automatic Film Reader
Brad King	WSMR Instrumentation Div.	Automatic Film Reader