Air Force Institute of Technology

AFIT Scholar

3-1998

Real Time Detection of Anomalous Satellite Behavior Has Ground Based Telescope Images

Geoffrey S. Maron
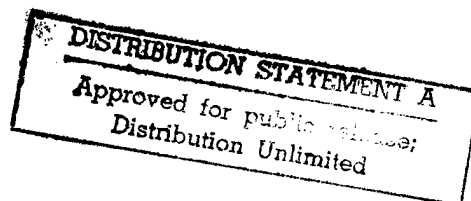
Follow this and additional works at: https://scholar.afit.edu/etd

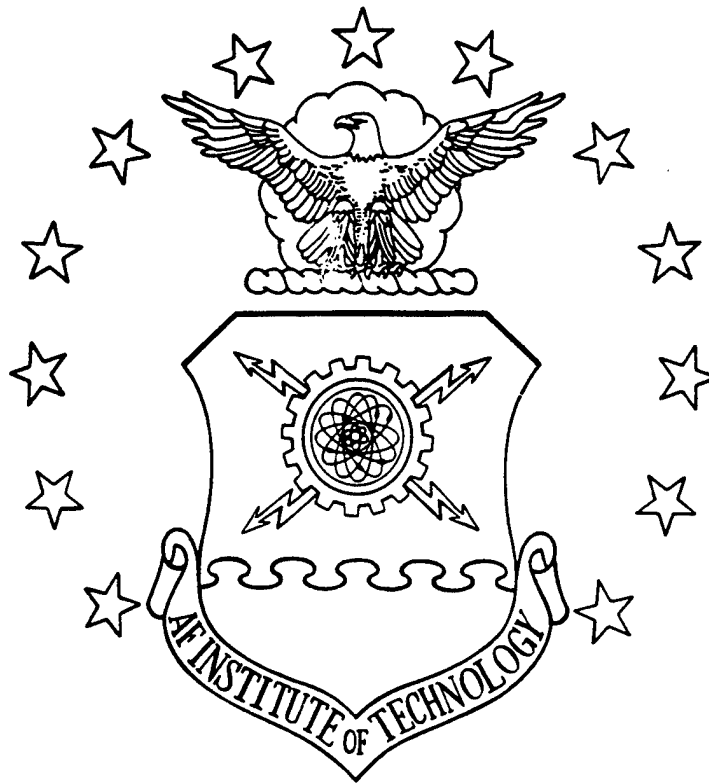Part of the Signal Processing Commons

Recommended Citation

19980519 088

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GOA/ENG/98M-01

Real Time Detection
Of Anomalous Satellite Behavior
From Ground-Based Telescope Images

THESIS

Geoffrey S. Maron
Lieutenant, USAF

AFIT/GOA/ENG/98M-01

Approved for public release; distribution unlimited

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

Real Time Detection

Of Anomalous Satellite Behavior

From Ground-Based Telescope Images

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Operations Research

Geoffrey S. Maron, B.S.

Lieutenant, USAF

March 1998

AFIT/GOA/ENG/98M-01

Real Time Detection

Of Anomalous Satellite Behavior

From Ground-Based Telescope Images

Geoffrey S. Maron
Lieutenant, USAF

Approved:

_____   _12 MAR 98_
Chairman   date

_____   _12 MAR 98_
                                             date

Table of Contents

List of Figures

List of Tables

AFIT/GOA/ENG/98M-01

Abstract

Air Force analysts are faced with the task of monitoring satellites with ground-based telescopes. Images are collected and analyzed in a time-consuming and subjective effort to detect any behavior that is anomalous. This research maximizes use of *a priori* information to create an automated, real-time satellite behavior classification tool.

Using modeling software and knowledge of a satellite's orbit, reference imagery is created for each measured image in a satellite pass. Features are extracted from the measured and reference image pairs that provide good overall gaussian classification accuracy (85%), reduce the dimensionality of the problem (from 32,768 down to 3), and are least dependent on data partitioning. The statistical image pair classifier is tested for robustness to atmospheric distortion, and training data requirements are explored.

Satellite behavior is classified by counting the classification results for the image pairs in a satellite pass. A binomial analysis of the classification technique predicts virtually 100% classification accuracy of satellite behavior. This research demonstrates the validity of model based satellite behavior analysis.

Real Time Detection

Of Anomalous Satellite Behavior

From Ground-Based Telescope Images

1. Introduction

1.1 Explanation of the Problem

Air Force personnel are involved in the task of Space Object Identification (SOI). This task includes a variety of activities with the purpose of understanding and recording the characteristics of the numerous entities orbiting our planet. The number of space objects is increasing at a growing rate, requiring greater amounts of effort to maintain current information on each object of interest. As the number of space objects increases, the need for automated analysis tools becomes more acute, especially in this era, when a growing number of the possible threats to our nation's security are space-based.

While numerous software packages are available that help track and image satellites in orbit, the image analysis task is performed by trained analysts. The analyst is faced with a series of images from a satellite's pass taken with a ground-based telescope. Because information is maintained on each and every orbiting satellite, the images are taken of a known satellite, in a known orbit. The analyst must determine if the imaged satellite is behaving in a new or anomalous manner of strategic interest.

This analysis is an important, time consuming, and subjective portion of the SOI task. An automated tool capable of objective, real time analysis would ensure consistent, timely knowledge of satellite behavior. With this knowledge readily available, the Air Force would be in a better position to ensure the security of the nation.

1.1.1 Problem Statement. Explore pattern recognition techniques based on a full utilization of available information for use in anomalous satellite behavior detection algorithms. Specifically, construct a computer algorithm capable of classifying a

1

satellite's behavior as normal or anomalous based on a collection of telescope images from a single satellite pass.

1.2 Approach

The problem as stated in Section 1.1.1 includes the phrase "full utilization of available information" due to a review of previous work done on the SOI problem at AFIT. Initially, the intent of this thesis was to improve the performance of the techniques that had already been applied to this problem. However, while exploring previous work, a need to re-look at the problem itself became apparent. Work up to this point had addressed the images without regard to the information that is known about the image. For example, we know what satellite is in the images, and what that satellite is supposed to be doing. This type of information is known, and should be used. To solve this problem most efficiently, it is essential that the problem, including all relevant information, be thoroughly understood and exploited.

1.3 Scope

1.3.1 Problem Exploration. Only sets of images from two actual passes of a satellite are available for use in defining the problem. This data, as well as an understanding of the method by which it is obtained, aids in the process of creating simulated data sets large enough to properly test pattern recognition techniques. Section 2.3 provides a full description of the real data and problem.

1.3.2 Simulating Real Data. As only two sets of real data are available, this thesis must create and use simulated data based on knowledge of the real data. The initial simulated images were created using the SatTools software [9], but are changed significantly for use in this thesis. Simulated data are designed to be similar to the data that would actually be measured and used in the real world.

1.3.3 Selecting Features. Comparing the object within images of different origins, quality, and resolution requires a representation of the object that is immune to differences of these quantities between images. To compare objects, it is necessary to

extract features from the sets of images that ignore differences in creation source, and simply represent the orientation and shape of the satellite itself. Feature extraction can also reduce the dimensionality of the information used.

1.3.4 Pattern Recognition Algorithm. There are many types of pattern recognition tools available, with a statistical classification method usually being the baseline. The solution methodology in this thesis utilizes two classification steps. In the first step, each image within a pass is classified as normal or anomalous. Then, the satellite pass is classified as normal or anomalous based on the cumulative statistics of the image classifications.

1.4 Thesis Organization

Chapter 2 covers the background information of earlier research on this problem, as well as the basics of traditional pattern recognition. Chapter 3 details the solution process, the methodology used in creating simulated data, the feature extraction process, the pattern recognition techniques applied to the data, and the experiments performed. Chapter 4 reviews the results of the experiments conducted and discusses the meaning of those results. The final chapter provides conclusions and addresses the operational applications of this research.

2. Background

This chapter provides the background upon which this thesis is built. Initial investigation into the work done earlier on this problem prompted a review of the problem itself. This chapter will overview the work done by two previous AFIT students, Gary Brandstrom and Neal Bruegger, then delve into the problem. The chapter also contains basic information on pattern recognition and the Fourier transform.

2.1 Pattern Recognition

While the field of pattern recognition is full of complex terminology and algorithms, at the most basic level, it is simply an attempt to determine to which group an unknown entity belongs. In many problems, such as this one, the number of possible groups is two, but the number of possible groups is not restricted. In this thesis, the two categories are "normal" and "anomalous" in reference to satellite behavior [1].

To make most pattern recognition problems manageable, pre-processing in the form of feature extraction is usually required. Features must be carefully devised so that differences in the features can indicate class membership with the chosen pattern recognition algorithm. Features are usually required to reduce the dimensionality of the data, as well as filter the data for the information that provides a class distinction [1]. In this thesis, two 128x128 pixel gray scale images are the initial data from which features must be extracted. Full storage of the information in the two images requires 32768 specific values. By extracting features from the images, it is possible to reduce the number of values that represent the information in the images to under 100, while still allowing successful classification.

Once features have been collected, pattern recognition algorithms require training. Training can be either unsupervised, or supervised, with the difference being the application of knowledge about the class membership of the training data. This thesis is concerned only with supervised training. The simplest algorithms for pattern recognition are based on statistics. A collection of training data can be used to determine a probability density function with respect to the features for each class. When a new set of

features must be classified, it is possible to assign a probability of class membership to that set of features. In this manner, classification is accomplished [2]. Many additional algorithms have been invented for use in pattern recognition. The previous work done in the SOI area at AFIT evaluated the performance of two different pattern recognition algorithms in the SOI problem.

2.2 Previous Work

Work on this problem began at AFIT with Captain Gary Brandstrom (GSO-95D). His thesis explored two different spatio-temporal pattern recognition techniques, Hidden Markov (HMM) models and the Feature Space Trajectory Neural Network (FSTNN), in an effort to determine which would be better in solving the SOI problem [10]. Captain Neal Bruegger (GOR 97M) worked to improve the performance of the FSTNN by improving the temporal aspect of the algorithm [7].

2.2.1 Problem as Approached by Brandstrom and Bruegger. Brandstrom and Bruegger approached the SOI problem with the statistical and nearest neighbor type classifier because of the infinite possible viewing angles and the fact that for each pass of the same satellite, the images taken of that sequence can appear quite different. These uncertainties would allow no knowledge of what the satellite should look like, and thus no known image sequence against which to compare the actual image sequence produced by the telescope.

2.2.2 Overview of Brandstrom Accomplishments. Brandstrom created a simulated data set with which to accomplish his work. Using the SatTools software package, he created a data set consisting of image sequences made up of 20 images taken with either 8 or 10 seconds between each image. All the images are of the same satellite. Brandstrom used this simulated data to make comparisons between two algorithms: the HMM and FSTNN. He found the FSTNN worked better than the HMM.

2.2.2.1 Hidden Markov Model. The HMM is a statistical model often applied to speech recognition problems that determines an overall probability that a sequence of observations (images) belongs to a certain class. They have been shown in work done by

5

Fielding and Ruck to be capable of classifying images of 3D objects based on the way the features and viewing angle for each image change with respect to time [3]. The overall probability of belonging to a certain class is determined by the probabilities of state to state transitions, where each state represents an observation at a certain time step [4].

2.2.2.2 Feature Space Trajectory Neural Network. The FSTNN was initially developed by Neiberg and Casasent to estimate the pose of an object in an image [5,6]. The training stage of the FSTNN consists of storing the characteristics, or features, for a set of observations. These observations should span the range of the possible poses of the object. A linear feature space trajectory is then assumed to connect the observations within the feature space. When a new image, or set of features, is available for test, the algorithm makes a perpendicular projection onto the nearest segment of the feature space trajectory. The value for the pose is determined by interpolation along that portion of the FST (Figure 1).
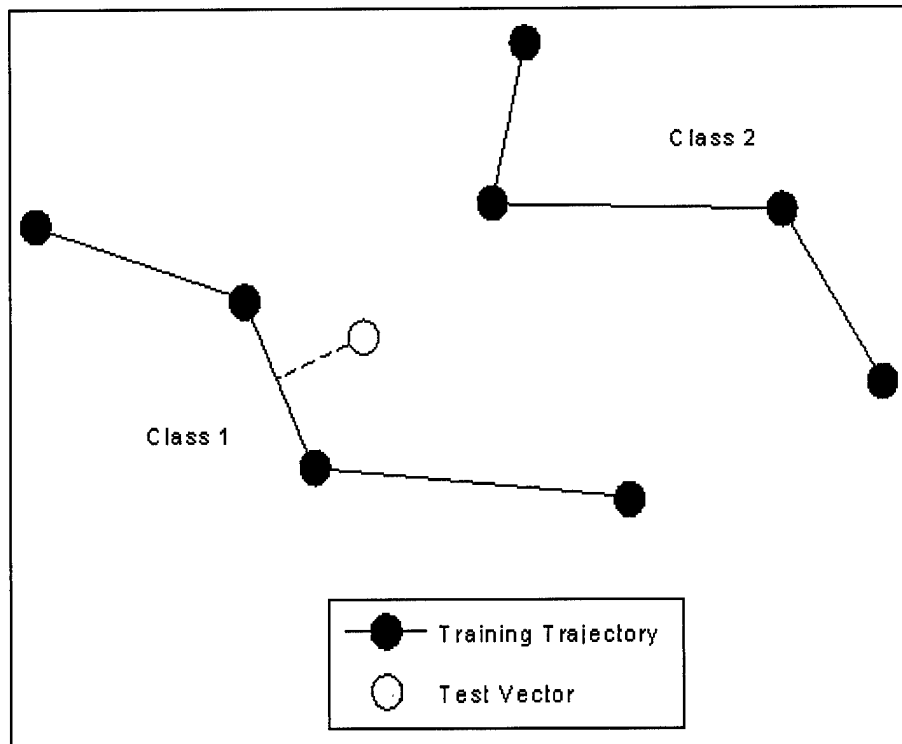
Figure 1. FSTNN as applied to pose estimation

If we consider the set of observations as not only pose observations, but temporal observations as well, we have a tool with which temporal information can be taken into account during the classification process. We could use a perpendicular projection of a new point in feature space to determine both the pose and the instant in time represented by the new point. The feature space trajectory becomes a trajectory through feature space as well as through time, which can help to narrow the problem solution space somewhat.

While the FSTNN was initially designed to answer pose estimation questions, it can also be used as a template matching type of classifier. In the simplest classification problem, we have two separate classes represented by two separate feature space trajectories. In order to classify a new sample, we simply create a feature space trajectory representation of the sample, and then compare the sample feature space trajectory against the feature space trajectories representing the two separate classes. With a distance measure, we can compute a measure of how close the sample lies to each of the class feature space trajectories. Class membership is chosen to correspond with the smaller distance.

Temporal information is inherent in the feature space trajectory when the observations that make up the points in that features space trajectory are sequenced through time. When using a FST as a template against which to compare the feature space trajectory of a new sample, temporal information is not considered. Attempting to use the temporal information was the purpose of Neal Bruegger's thesis work.

2.2.3 Overview of Bruegger Accomplishments. Bruegger continued the work that Brandstrom had started. He took the FSTNN and incorporated time into the algorithm, forcing the algorithm to consider the time sequence of observations within the feature space. Previously, the algorithm had been time insensitive, even though time was an important known bit of information. He proposed, and tested, two separate methods for incorporating time into the FSTNN: Dynamic Time Warping and Uniform Time Warping. Both methods of considering time in the anomalous / normal decision process were capable of improving performance.

2.2.3.1 Dynamic Time Warping (DTW). Dynamic Time Warping is one method for incorporating time into the FSTNN. DTW is based on the knowledge that the observations in the feature space are actually also observations in time. When using a FSTNN as a template against which to match an unknown trajectory, the DTW algorithm restricts the segment against which a perpendicular projection can be made. By restricting the segments available for projection, the algorithm forces the proper motion through feature space.

2.2.3.2 Uniform Time Warping (UTW). Uniform Time Warping is a less flexible method of dealing with time than DTW. UTW considers the test and training feature space trajectories as continuous transitions of the same overall duration. The UTW algorithm segments both the test and training trajectories into a given number of equidistant points. Each equidistant point on the test trajectory corresponds to a single point on the training trajectory. The distance between each corresponding point is summed for the final distance measure. This method requires equal duration for both the test and training measurements. [7]

2.3 Fourier Space

As the Fourier space representation of images will be used in this thesis, this brief section on the interpretation of that space is included. While the mathematics behind the Fourier transform is somewhat complex, the interpretation of the transform can be understood without it. In one dimension, the Fourier transform represents a signal as a weighted sum of sinusoids at different frequencies. These sinusoids, when added together will recreate the original signal. Thus, the collection of frequencies and amplitude at each frequency is a representation of the original signal in the frequency domain [8].

In two dimensions, the same interpretation applies. The frequency domain representation of a two-dimensional waveform is a collection of sinusoids at particular frequencies, amplitudes, and phases. These frequencies, amplitudes and phases represent the sinusoids required to reconstruct the original waveform. When a two-dimensional Fourier transform is applied to an image, the resulting Fourier representation of the image

is a matrix of values the same size as the original image. Each member of the matrix contains a real and imaginary portion. From this imaginary number, both energy and phase can be calculated.

2.4 Problem Description

It should not be surprising that Bruegger's work incorporating time into the solution algorithm improved the results. The more information that is known about a system, the easier it should be to solve for the unknowns. Bruegger's addition of sequence information to the algorithm was a first step in a more complete look at the problem. One purpose of this thesis is to discover the other parts of the problem that earlier algorithms ignored to provide an improved method of determining if the satellites being imaged are behaving normally or abnormally.

2.4.1 Actual situation.
The Air Force Maui Optical Station (AMOS) images 8-10 different satellites with a ground-based telescope that are of interest to the Air Force in relation to this thesis. Every orbiting object of interest is tracked by the Air Force, so information regarding the identity and position of any specific satellite is *a priori* knowledge. Using the orbital information for a known satellite, the AMOS telescope can image an entire pass. The data from a single pass can usually be reduced into somewhere between 3 and 60 good images of the satellite. Each reduced image is actually made up of numerous images that have been processed together to produce a single good image. The reduction process is a method of combating the problems of imaging through a turbulent atmosphere.

There is no periodicity to the good images within a pass. The images may not be evenly temporally spaced within a pass, but the point in time for each image is known. Along with a time stamp, the telescope pointing angle and angle of elevation for each image are also recorded. Based on this available information, the position of the satellite in the sky is precisely known.

The atmospheric conditions are the only true unknowns in the image acquisition process. The quality of each image is strongly dependent on weather conditions, atmospheric turbulence, and the fact that most of the images are taken during daylight

9

hours. The variability in these conditions results in degraded images, which makes it difficult to visually determine if the satellite is behaving properly.

One more very important piece of information is also available for use in this thesis: the proper appearance of the satellite given its position with respect to AMOS is known. This knowledge is a key element in solving this problem, but has been ignored in previous work. Mathematical models of each satellite in the Air Force inventory exist which can be used to create simulated imagery of any satellite given any viewing angle and viewing location. RDCSIM (Research and Development Consortium Simulation) is a satellite imaging simulation that is currently used to support the efforts of the AMOS/MHPCC (Maui High Performance Computing Center) Research and Development Consortium. RDCSIM is heavily based on the simulation software developed for the Phillips Lab called SATSIG (Satellite Signature). Both these tools are used to create simulated imagery of satellites for which mathematical models exist [9].

2.4.2 Actual images. Only two sets of actual images of a known satellite are available for this thesis work. This mandates and aids the creation of simulated data. The actual images are from passes of a TD81 satellite taken on 12 and 15 October, 1996 (Figure 2).

Figure 2. Pristine image of TD81 satellite.

The 12 October pass includes only 6 images of medium quality. The 15 October pass contains 48 images of slightly better quality. The satellite can be easily recognized in many of the images in the 15 October pass (Figure 3).
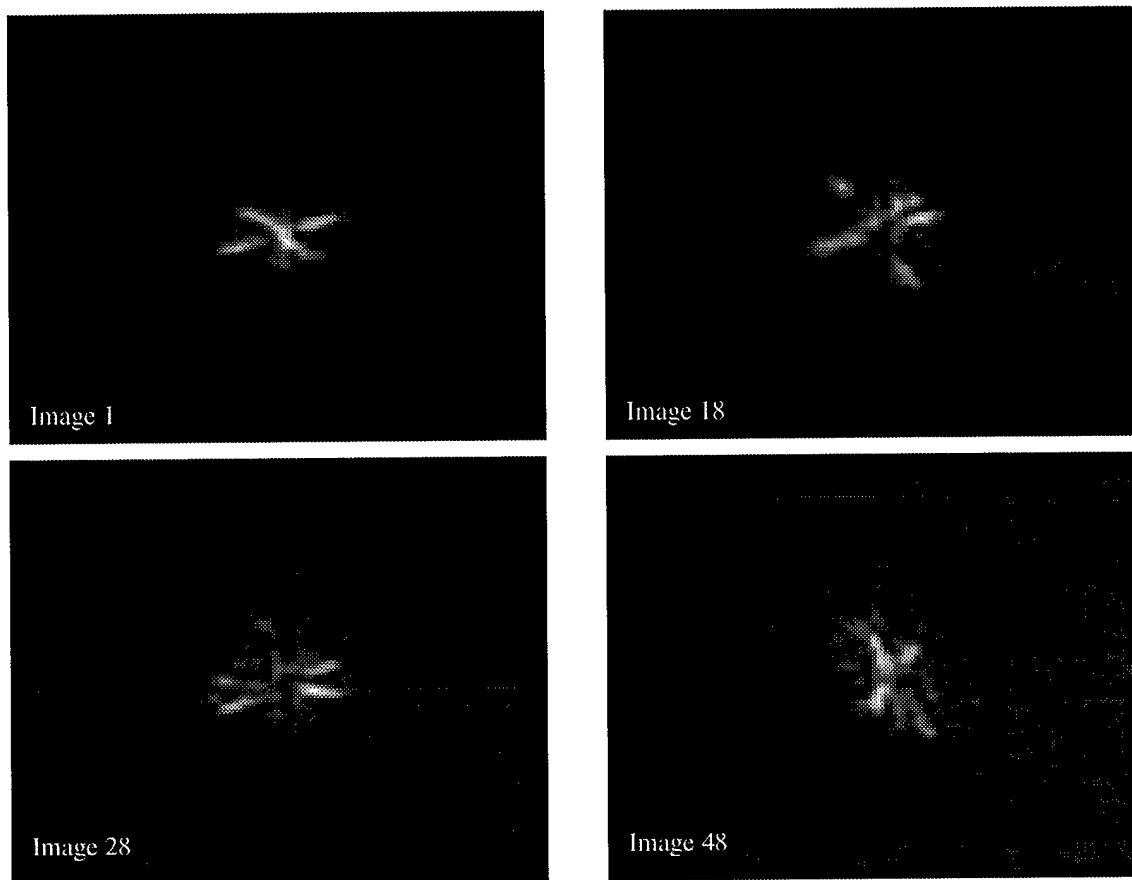
Figure 3. Samples of good images from 15 Oct 96 pass.

While the shape of the satellite is easy to see, most of the details are not. Without the details, it is difficult to determine the orientation of the satellite from a single two-dimensional image. The body of the satellite could be pointing towards or away from the telescope, but without details, the image will appear as a simple cross shape (Figure 3, Image 1). Only when the body and solar panels are not lined up with respect to the line of site of the telescope, does the satellite appear to be more than a simple cross (Figure 3, Image 18). To add to the complexity of the problem, image quality is not consistent throughout a pass. The same October 15 pass that contains the images in Figure 3 also contains the images in Figure 4.
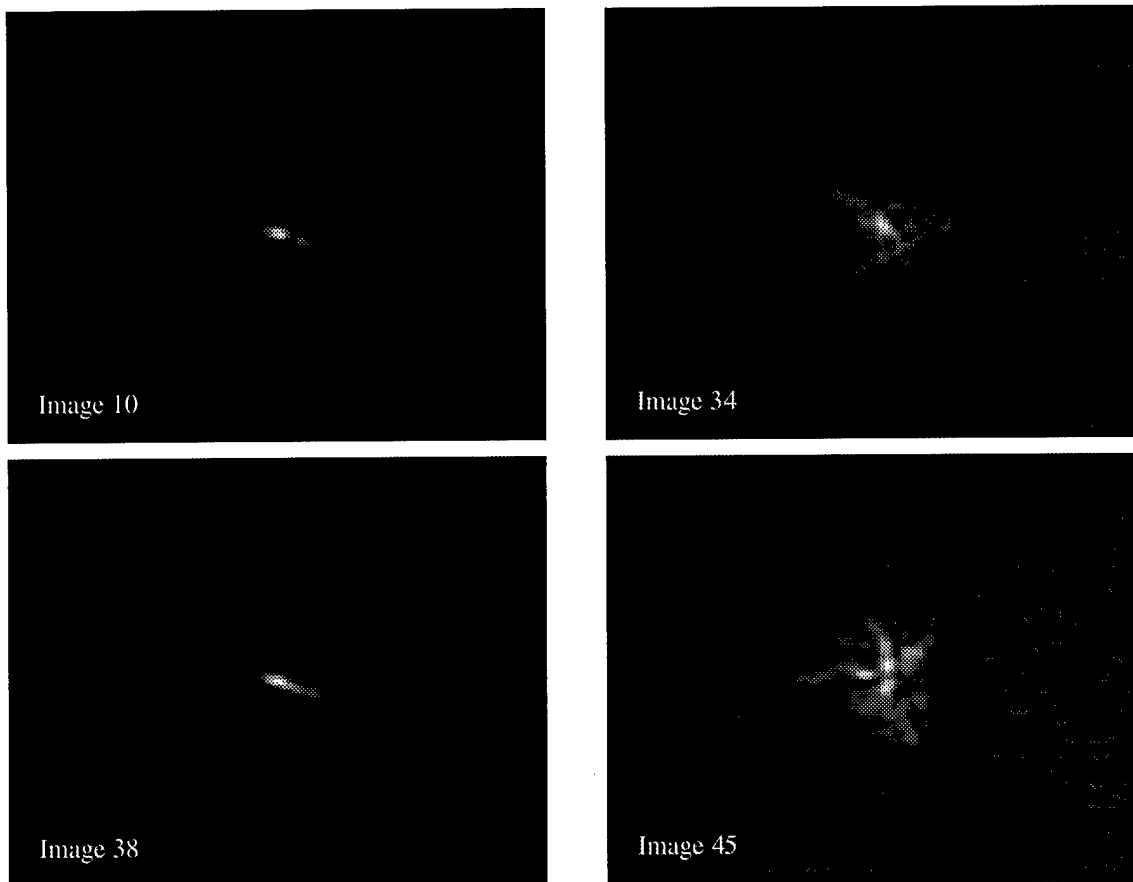
Figure 4. Sample of bad images from 15 Oct 96 pass.

Examination of the small amount of available real data aids in construction of simulated data by revealing the quality of imagery that can be expected. Of particular interest are issues such as satellite image size within the frame, intensity of satellite to background pixels, visible detail, and image blur. By concentrating on these issues, it should be possible to make simulated data that approaches the operational situation. After exploring and understanding the background information, the path to a solution should be easier to conceive and follow.

3. *Methodology*

In this chapter a solution process is proposed. The process must be tested by experimentation on manufactured data. This chapter will first provide an overview of the solution procedure, then delve more deeply into the process used for creating the data for experimentation. Portions of the solution procedure will be detailed, and, finally, the experiments described.

3.1 Solution Procedure

By delving more deeply into the problem and understanding the information that is available, a very basic solution methodology can be seen. The *a priori* information regarding a satellite's orbit in conjunction with the information corresponding to each image can be used to create a simulated reference image of how the satellite should appear in each image. This computer generated image can be compared against the measured image to determine how the satellite appears to be behaving at that moment. The results of the comparisons between image pairs for the entire pass can be examined to determine if the satellite is behaving normally (Figure 5).

Figure 5. Solution Procedure

3.2 Database

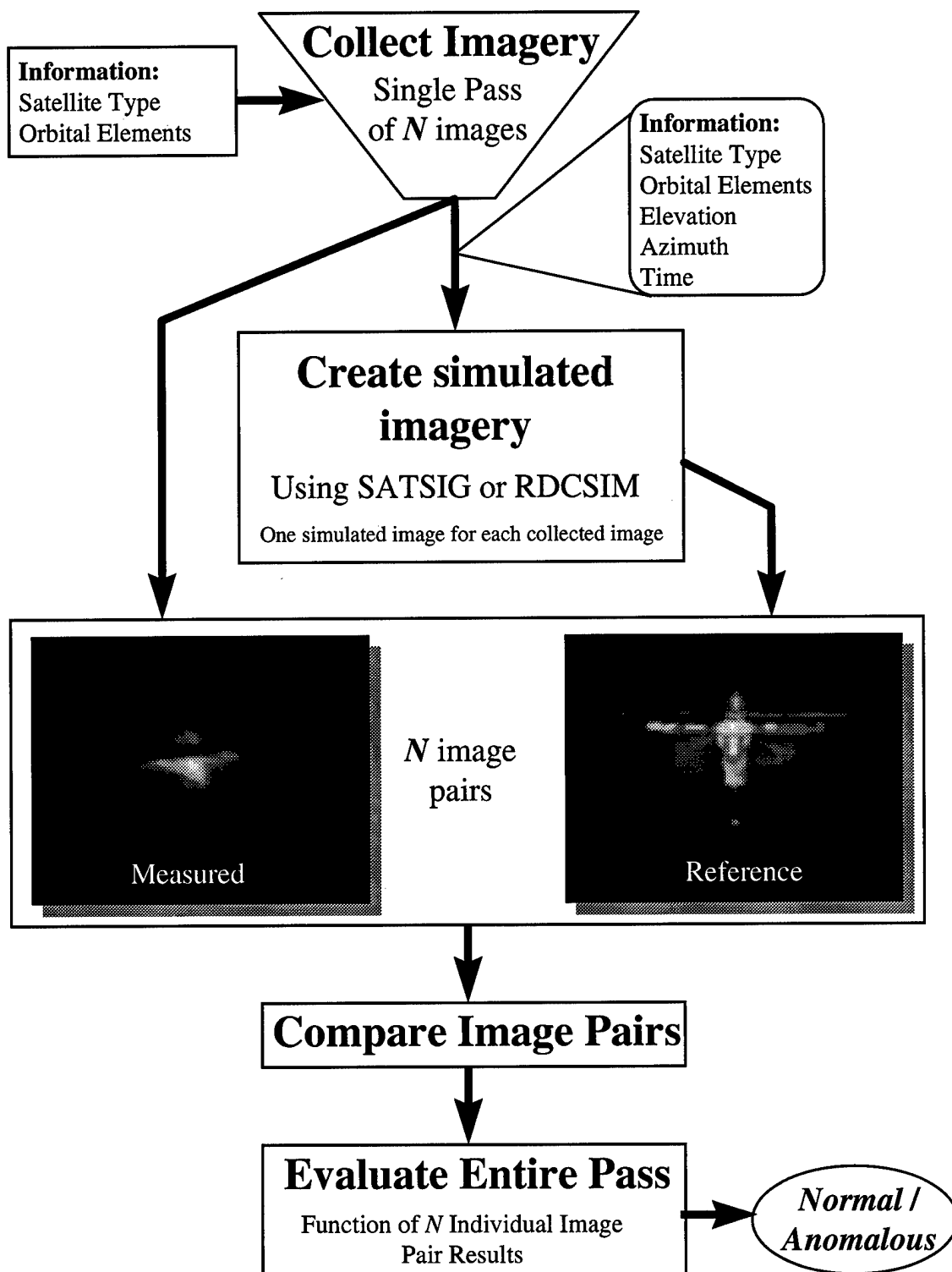Due to the lack of sufficient amounts of measured data, this thesis requires the synthesis of simulated data. Pseudo-measured images will be created by mimicking atmospheric effects. Once the data has been created, it will be assumed to be "measured" data in this thesis. The simulated database consists of image pairs similar to the ones that would exist before the *compare image pairs* step in the solution procedure (Figure 5). Each image pair will contain a pseudo-measured and a simulated reference image, where the pseudo-measured image is also a simulated image.

A portion of the pristine TD81 imagery initially created by Brandstrom will be used as the base imagery for the data created in this thesis. In total, this thesis makes use of 120 satellite passes. Each pass contains 20 images for a total of 2400 images. Approximately half of the passes were created as anomalous, but the approach in this thesis makes an anomalous pass one in which the pseudo-measured images do not match the reference imagery within each image pair. Thus, any of the passes can be made into either normal or anomalous passes, by simply creating an image pair that either matches or does not match. Figure 6 depicts the process of creating the simulated data for a normal pass with atmospheric distortion.

Anomalous passes are created in a fashion similar to the normal pass creation, differing only in the final image pairing. For the anomalous passes, the reference image from which the pseudo-measured image is created is not used as the reference image in the final pairing. Rather, a simulated image from another pass is used as the reference image. This gives anomalous passes with image pairs that may be similar, but do not match.
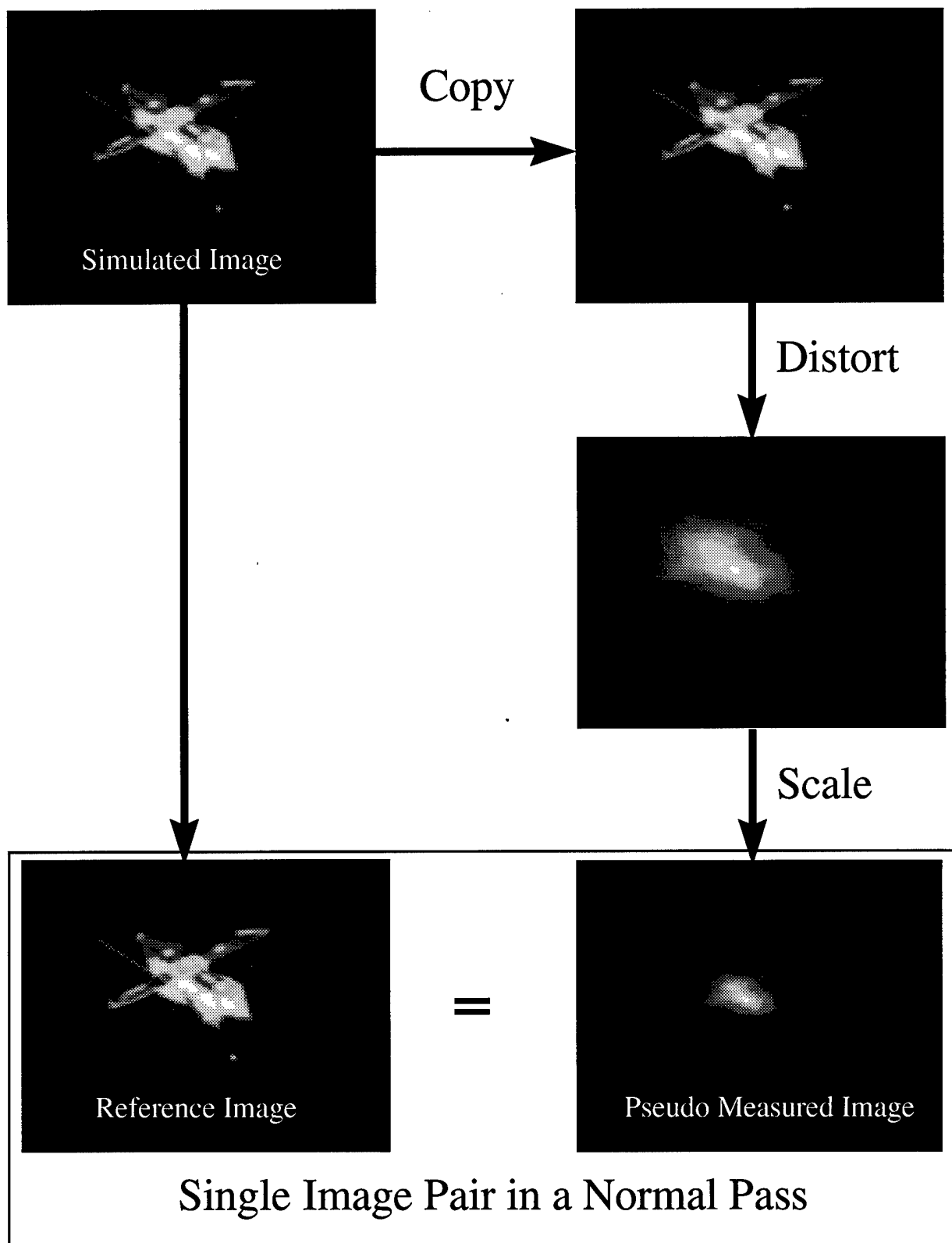
Figure 6. Process for creating normal pass image pairs.

Seven different data sets are used in this thesis. The first data set consists of 20 satellite passes with the pseudo-measured images all distorted at the same level. The remaining six sets of data are all based on the same 100 satellite passes, with the pseudo-measured images distorted to 6 different levels by means of different optical transfer functions (OTFs). OTFs are described below.

Table 1. Description and Optical Transfer Function identification of data sets.

Data Set	Image Pairs	Degradation Level	OTF
1	400	Medium	3
2	2000	Low	1
3	2000	Medium-Low	2
4	2000	Medium	3
5	2000	Medium-High	4
6	2000	High	5
7	2000	Random	1-14

3.2.1 Distortion. To create pseudo-measured images that approximate the quality of the real images, it is necessary to degrade the pristine simulated imagery in a manner similar to the degradation affected by the atmosphere and optics. Degradation of the images is accomplished by using an optical transfer function. The OTF is applied to the Fourier space representation of the image. The resulting representation is inverse Fourier transformed to obtain the degraded image (Figure 7). This process reduces the level of fine detail in a manner similar to the degradation affected by the atmosphere.
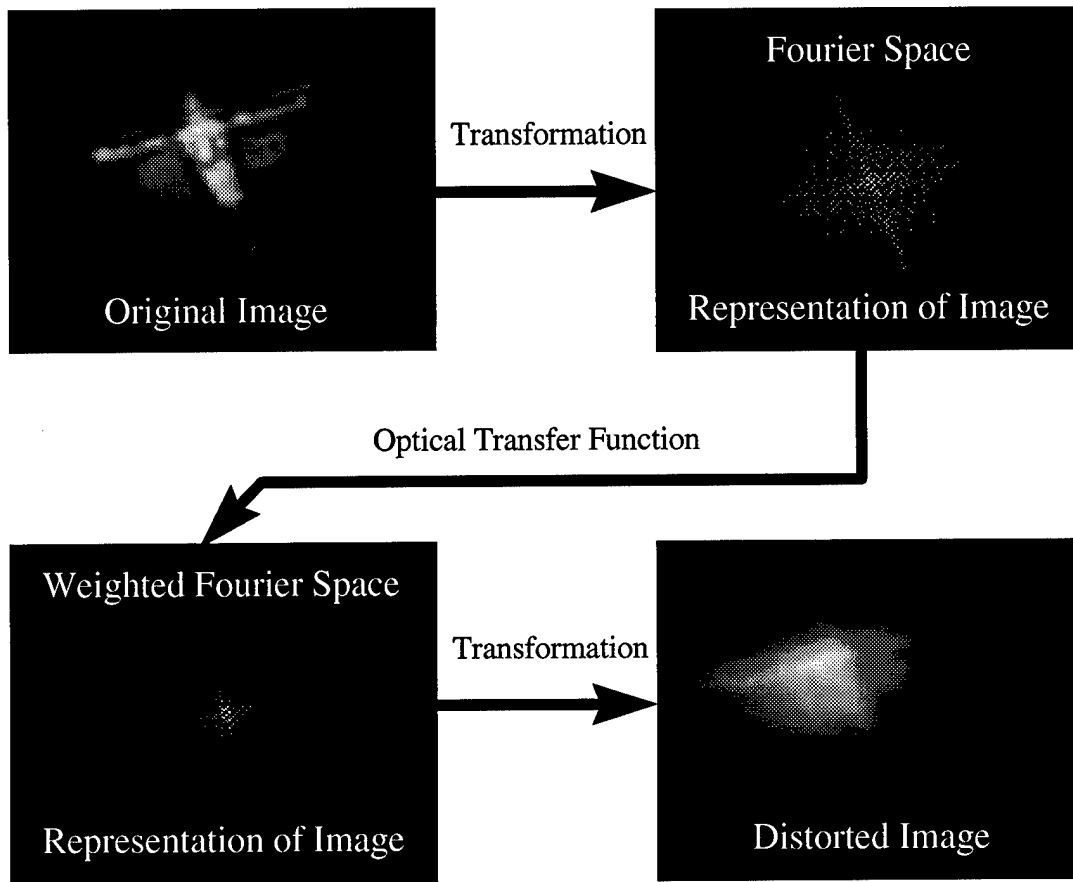
Figure 7. Process of distorting images.

An OTF can be calculated using the HYSIM5 program [7]. This program calculates an OTF corresponding to certain levels of atmospheric turbulence and lens distortion. The OTF is used as a matrix of weights for application against the Fourier space representation of the image. Figure 8 plots the magnitude of one OTF used in this thesis. The height in the Z-axis represents the weighting applied to the corresponding point in the Fourier space representation of the image.
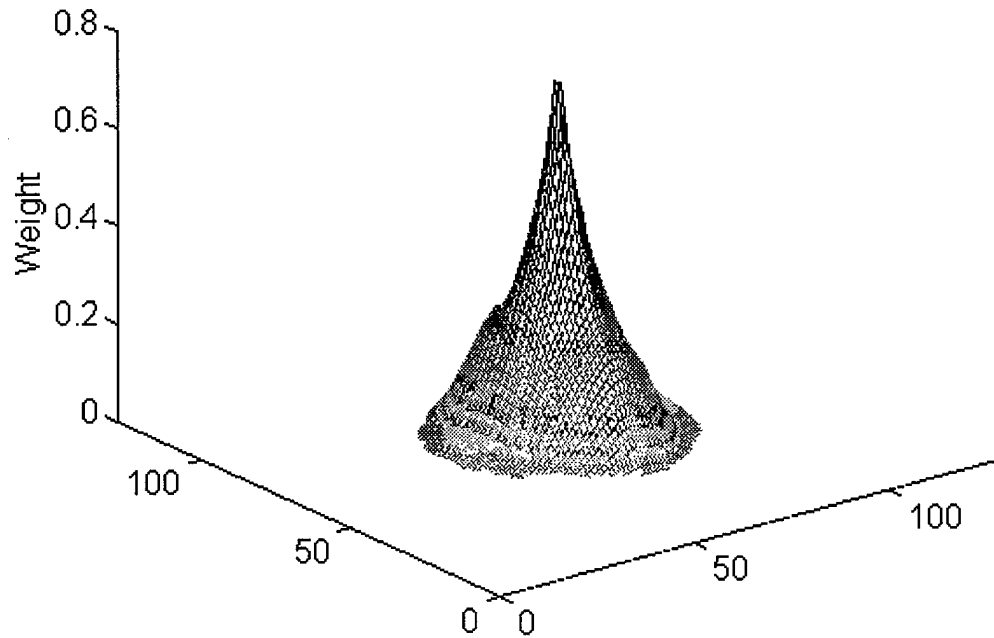
Figure 8. Optical Transfer Function used for image distortion.

Fourteen different OTFs are utilized in this thesis for creating the 7 data sets (Table 1). Figure 9 shows the distortion affected on a pristine satellite image by each of the 14 possible OTFs. Data set one is created with the medium level OTF. Data set 2 through 6 are created by using a single OTF across the entire data set. The 5 OTFs used in these data sets range from low to high distortion (Figure 10). A larger value of normalized energy corresponds to less distortion. For comparison purposes, contour plots of the low and high distortion OTFs are shown in Figure 11.

Normalized energy is the percentage of possible weighting for a 128x128 OTF. The maximum weighting energy at any pixel is 1.0, so a fully weighted, non-distorting OTF contains a total energy of 128^2 and a normalized energy of 100%.
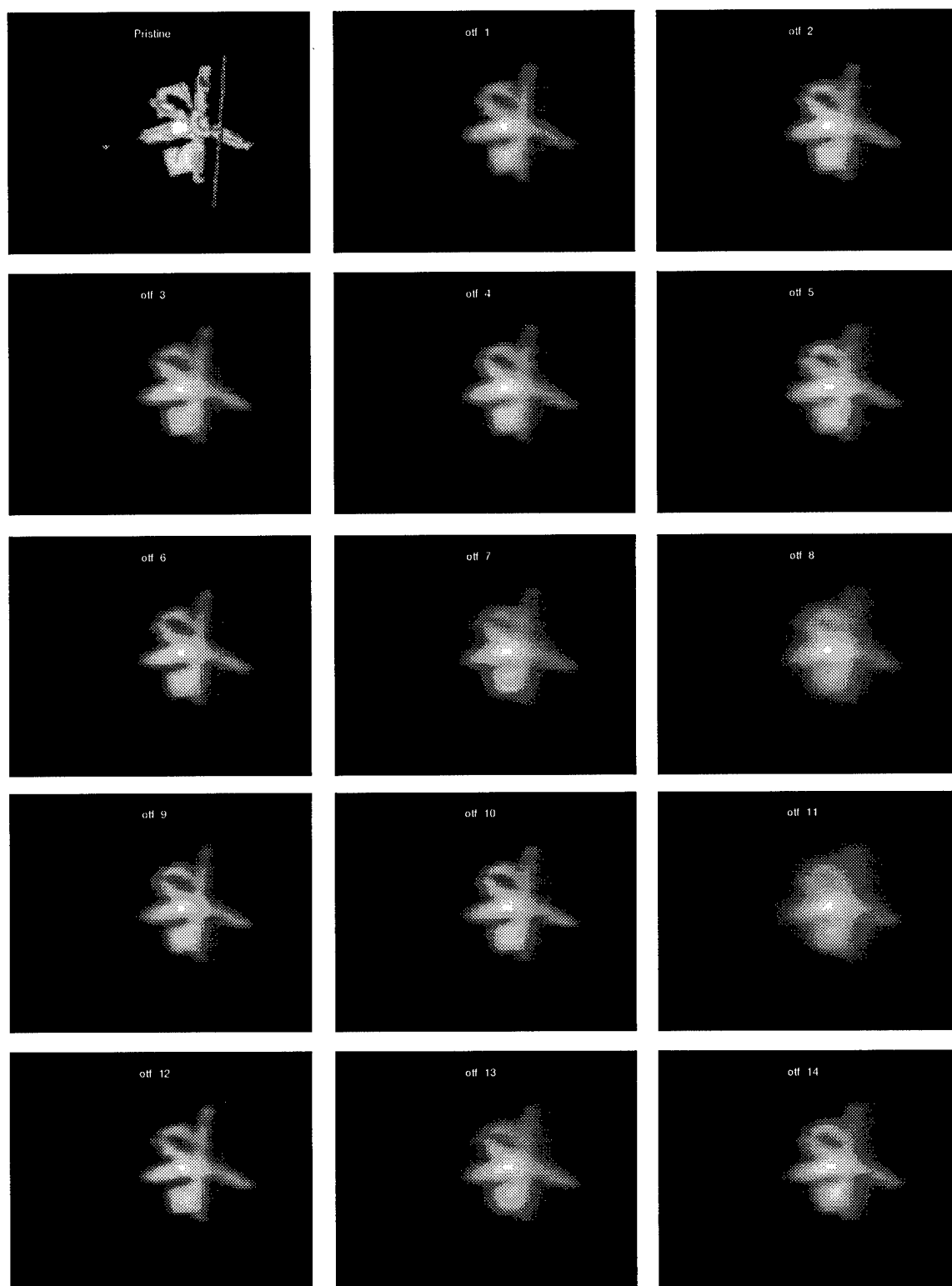
Figure 9. Distortion affected by each OTF.
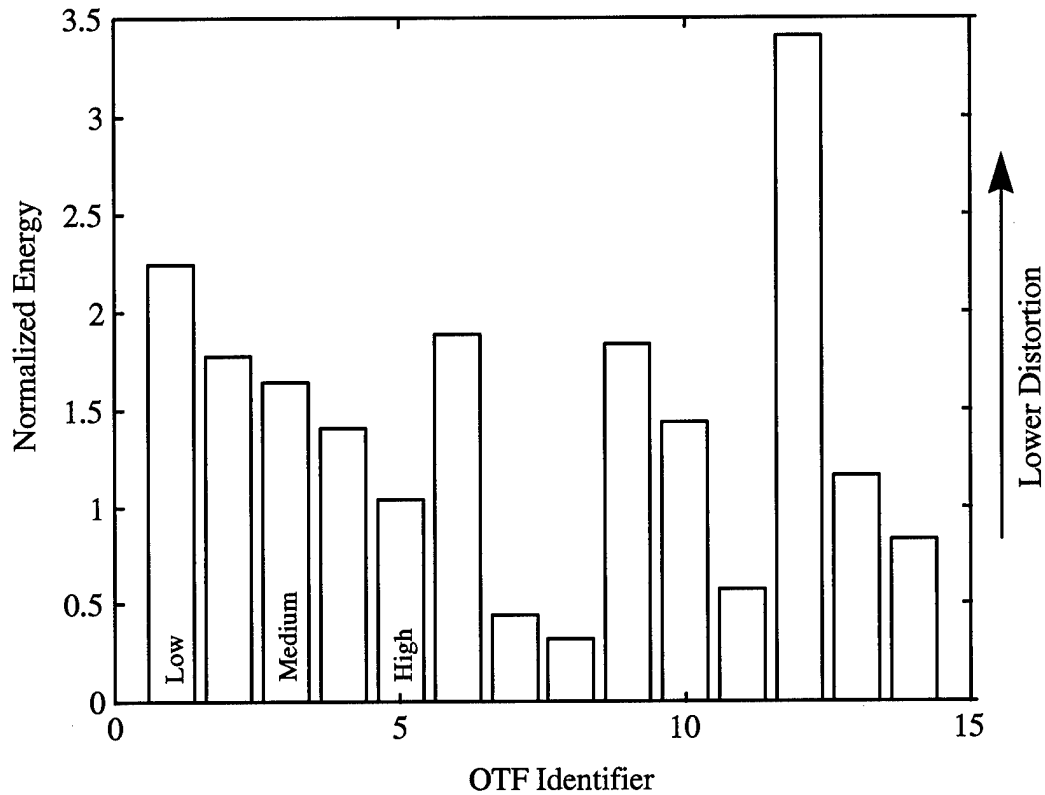
Figure 10. OTF normalized weighting energies.
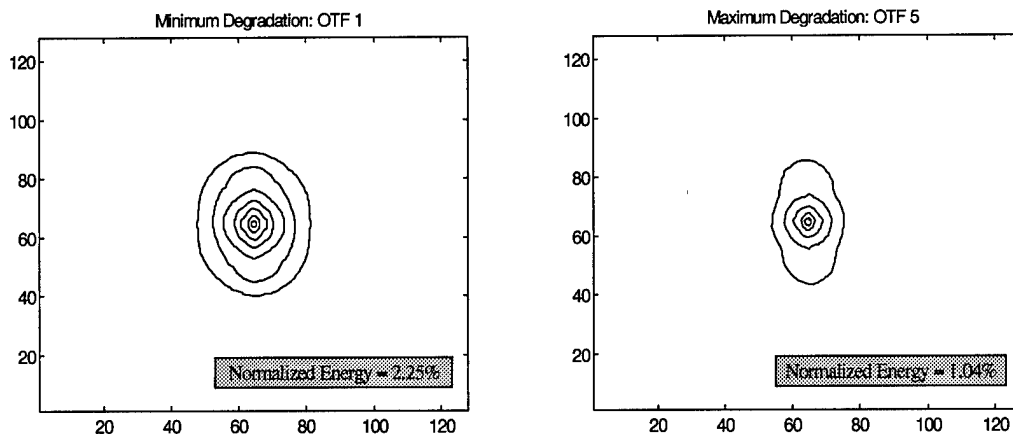
Figure 11. Selected OTF contour plots.

While data sets 1 through 6 are created with a single OTF applied throughout the data set, the final data set is created by randomly varying the OTF throughout each satellite pass. A varying OTF better approximates the true constantly changing

atmospheric distortion. In order to create the randomly varying data, five different OTF levels are randomly chosen from the possible 14 for each satellite pass. The OTF used for distortion of an image in a pass is a smoothing combination cycling through the 5 OTFs. Each image in a pass of 20 images is distorted with a different OTF. Images 1, 5, 10, 15, and 20 are distorted with the 5 OTFs randomly chosen from the possible 14. Image 2 is distorted with an OTF created as a combination of ¼ the image 1 OTF and ¾ the image 5 OTF. Image 3 is distorted with an OTF created as a combination of ½ the image 1 OTF and ½ the image 5 OTF. The remaining OTFs are created in a similar fashion to yield OTFs varying though time (Figure 12).
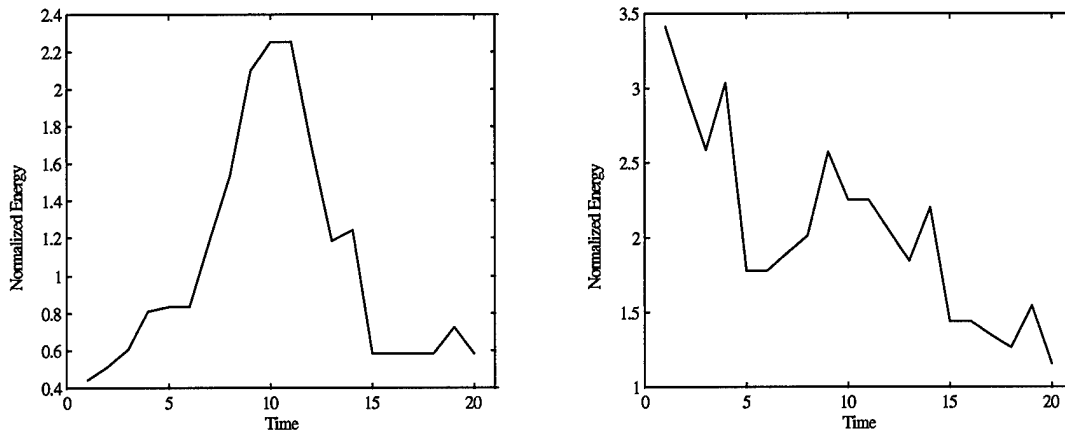
Figure 12. Normalized energy of sample random OTFs from data set 7 plotted against time. Time is represented by temporally spaced images 1 through 20.

Data sets 2 through 7 allow examination of the impact training and testing at different distortion levels can have.

22

3.2.2 Scale and Resolution. From analysis of the real imagery, it is apparent that the simulated images created by SATSIM are approximately twice the size of the images obtained by the telescope. To achieve the proper scale in the simulated data, the 128x128 images are first low pass filtered then scaled by one half. The smaller image must then be inserted into a 128x128 background image. The background image is simply a 128x128 matrix of zeros. Due to the smaller scale of the satellite within the image, information is lost because of lost information pixels, and the image pair comparison algorithm must be capable of working with scale differences through re-scaling or scale invariance.

3.2.3 Pairing. As shown in Figure 5, creation of image pairs with real data would be based on the proper orbital and telescope information. The information would be fed into the SATSIG program, which would produce a reference image corresponding to each real image. These two images would make up an image pair. Creating image pairs for the simulated data is the reverse process. A simulated image is used to produce a pseudo-measured image. These two images then make up the image pair for a normally behaving satellite. To create anomalous data, the images within each pair for an anomalous pass are chosen so they do not match.

Initially, 10 passes are created in which the satellites behavior is considered anomalous and another 10 passes in which the satellite behaves properly. Each pass in the first data set consists of 20 image pairs, for a total of 400 image pairs. Figure 13 and Figure 14 are examples of normal and abnormal satellite behavior. The image pairs in Figure 13 match, while the image pairs in Figure 14 do not match.

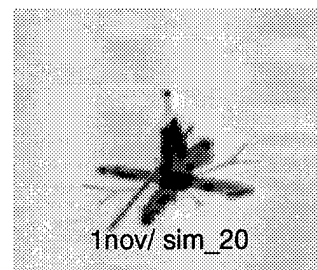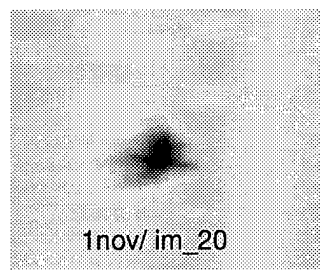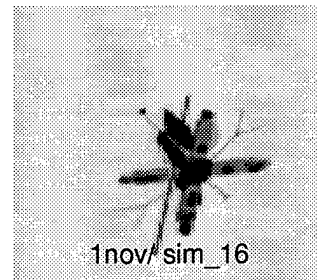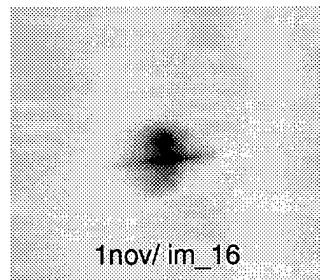Pseudo-Measured Image   Reference Image
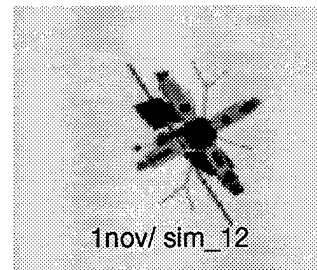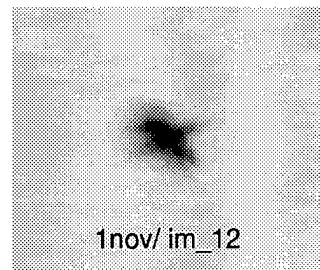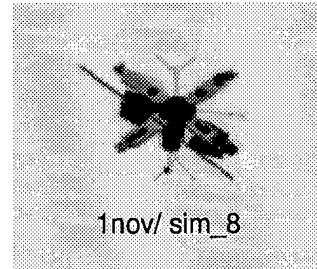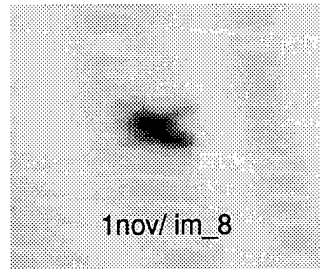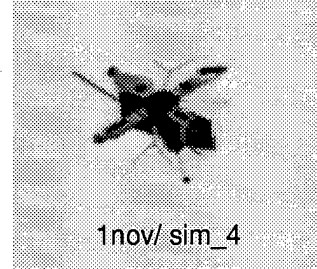
Figure 13. Selected image pairs of a pass with normal satellite behavior.

Pseudo-Measured Image   Reference Image
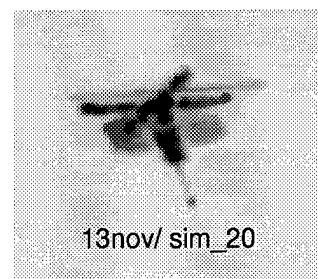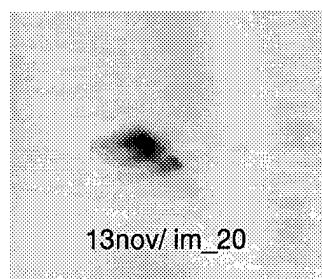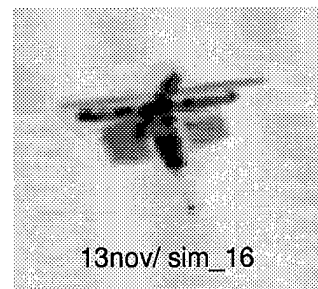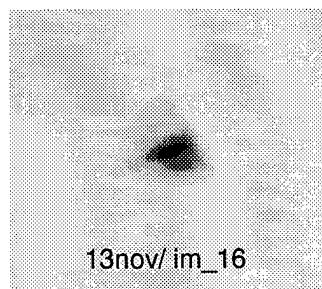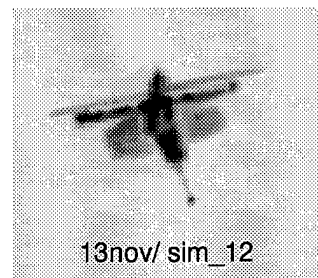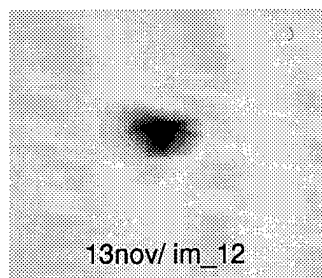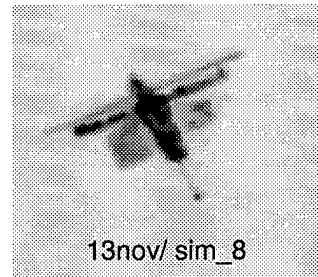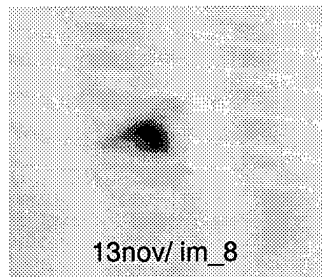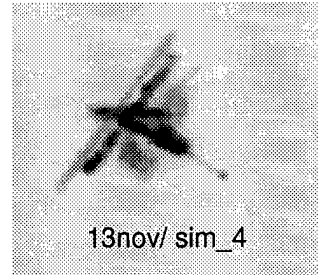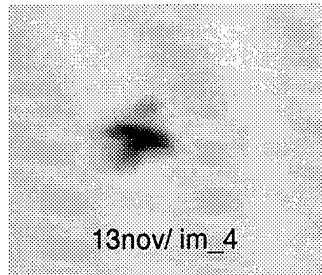
Figure 14. Selected image pairs of a pass with anomalous satellite behavior.

3.2.4 Data Set Use. Data set 1 is used to explore types of features and investigate initial classification techniques. Once initial exploration is complete, the remaining data sets are created. These data sets are based on 100 new satellite passes containing 20 images each. To explore the effect of differing atmospheric conditions, six sets of 2000 image pairs are created. The pseudo-measured images in each of the six sets are created using six different distortion levels. With these six sets of image pairs, it is possible to asses the impact on performance of training and testing with different OTFs. In total, data sets 2 through 6 contain 12,000 image pairs.

3.3 Data Representation with Features

With a database of simulated image pairs created, we must concentrate on the main step in the solution procedure, classification of these image pairs. The image pairs described in section 3.2 must be represented by a relatively small collection of features because of the nature of statistical classifiers. A statistical classifier is based on the class conditional probability density functions for each feature. Each PDF is estimated from data, and exponentially more data are needed for each additional feature. Feature extraction also reduces the dimensionality of the problem and can minimize classification-hampering extraneous information

To initially explore different types of features, a total of 26 different feature sets will be created. Each feature set is made up of a collection of 1 to 25 features. Feature sets will be combined into a feature vector to be used for classification of data set 1. Feature vectors will include two or more feature sets for a minimum feature vector length of two. The feature sets that prove promising will be further evaluated with data sets 2 through 7 to create a final feature vector that is capable of describing the image pair in enough detail to determine if the imaged satellite is behaving normally or abnormally.

3.3.1 Two-dimensional Fourier Space Features. The two-dimensional Fourier space provides a representation of an image that can be used to effectively capture shape and rotation information. Features derived from the Fourier matrix make up the bulk of the feature sets tested.

26

3.3.1.1 Block Features. Block features can be easily extracted from a Fourier matrix and are useful in describing the shape of an image. Each block feature is one specific value within a block defined to enclose the desired number of features. This block is generally close to the center of the Fourier matrix as the values close to the center of the matrix correspond to the more general shape parameters of the image. The block is not symmetric about the center of the matrix because the matrix itself is symmetric. The Fourier matrix does not, however, represent the image with half the number of values because, due to the nature of the Fourier transform, the block features have both a real and imaginary portion. The subset of 25 values extracted from the Fourier matrix are depicted in Figure 15.

Transform to
Fourier Space

Extract 25 Values as Block Features
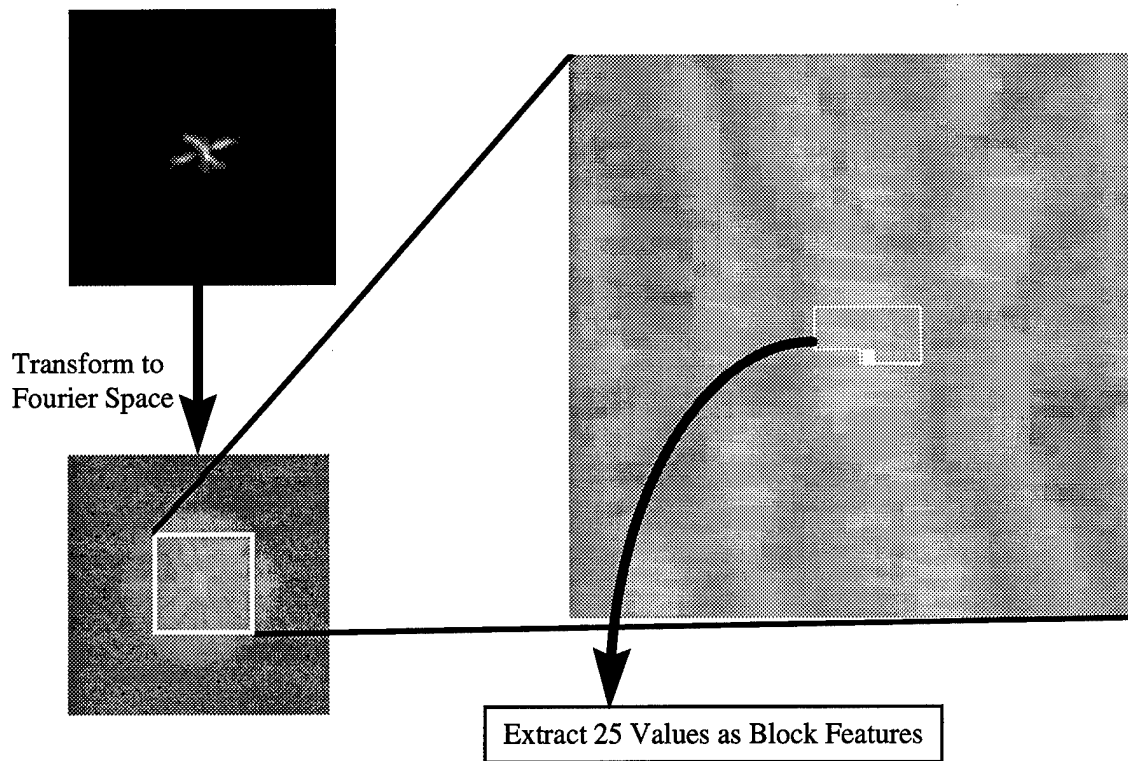
Figure 15. Fourier space block feature extraction.

The first two feature sets are defined to be the 25 block features for each image within an image pair. The issue of statistical normalization gives rise to two additional feature sets. Block features are often statistically normalized to ensure that the mean and variance of each feature are of the same magnitude [10]. For all features F_i ($i = 1...25$), a new feature F_i' is calculated:

$$F_i' = \frac{\overline{F} - F_i}{\sqrt{\text{VAR}(F_j)}} \quad j = 1...25$$

Two additional feature sets are constructed from the 25 normalized block features for each image in an image pair.

While initial exploration uses block features taken from the Fourier transform of the gray scale image, the features can also be extracted from the Fourier transform of a binary image created by thresholding (section 3.3.2.1). The block features in data sets 2 through 7 are extracted from the binary image.

3.3.1.2 Wedge Features. Rotational difference between images can be captured with wedge features. Wedge features are calculated by summing the energy within wedges of the Fourier matrix. Energy is defined as the complex value at each pixel times the complex conjugate of that value. Thus, energy at each spectral location is a real value. The summed energies from eight wedges of equal size are used as features. Wedges from only one half of the symmetric Fourier matrix are required. The process as well as an example of how the wedge features differ for rotationally distinct images are shown in Figure 16.
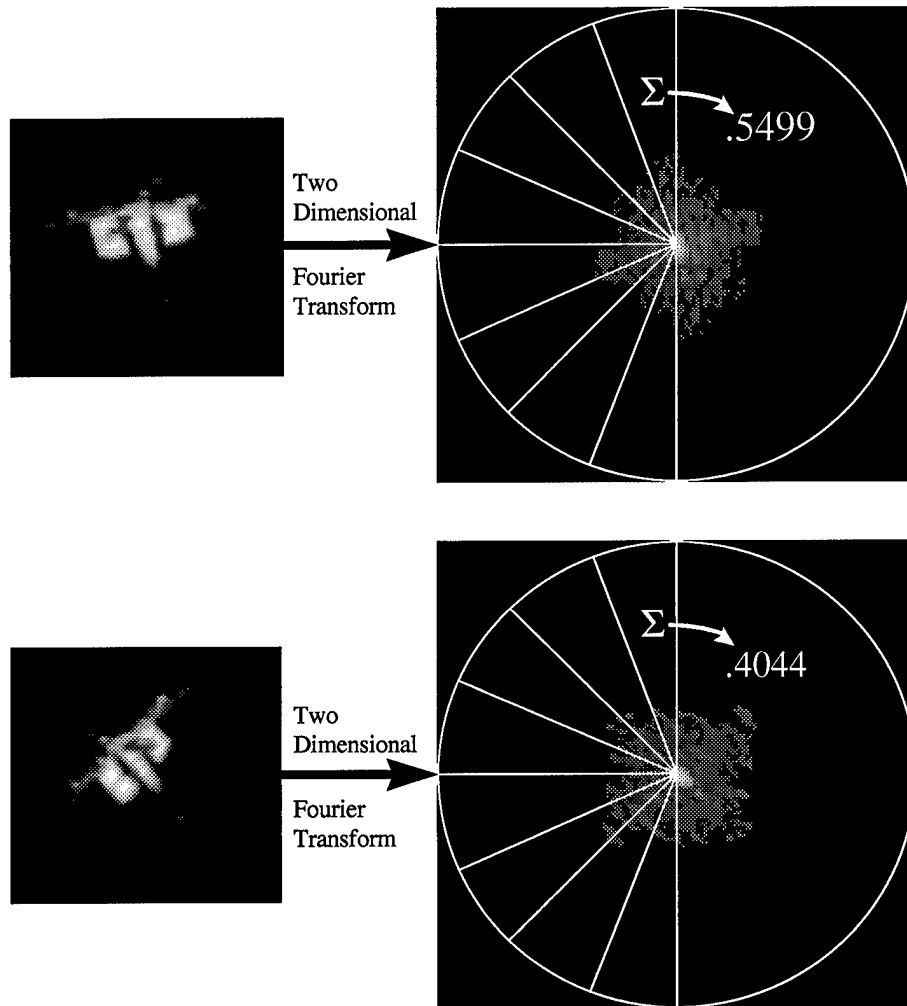
Figure 16. Fourier space wedge feature extraction with example wedge energy sums depicted.

Due to the size difference between the measured and reference portion of each image pair, it is important to have wedge features that are scale invariant. This is accomplished by energy normalizing the features. For all features G_i ($i = 1...8$), a new feature G_i' is calculated:

$$G_i' = \frac{G_i}{\sqrt{\sum_j G_j^2}} \quad j = 1...8$$

29

The two energy normalized wedge feature sets as well as a statistically normalized set of the same features add four additional sets for testing.

3.3.2 Moment-based Features. Moment-based features are another means by which the shape of an image can be represented. The moments of an image of two dimensions can be defined for both gray scale and binary images. To reduce the possible effects of background values on the moment calculations, the images will be transformed to binary images. This transformation should capture the satellite regardless of the background and make comparison between images with different background levels possible.

3.3.2.1 Conversion to Binary Images by Thresholding. To transform a gray scale image to a binary image, pixel values are compared to a chosen threshold and set to a binary value as:

$$f(x,y) \underset{0}{\overset{1}{\gtrless}} \text{Threshold}$$

For the real measured images, a threshold value of mean plus three times the deviation was empirically chosen (Figure 17).
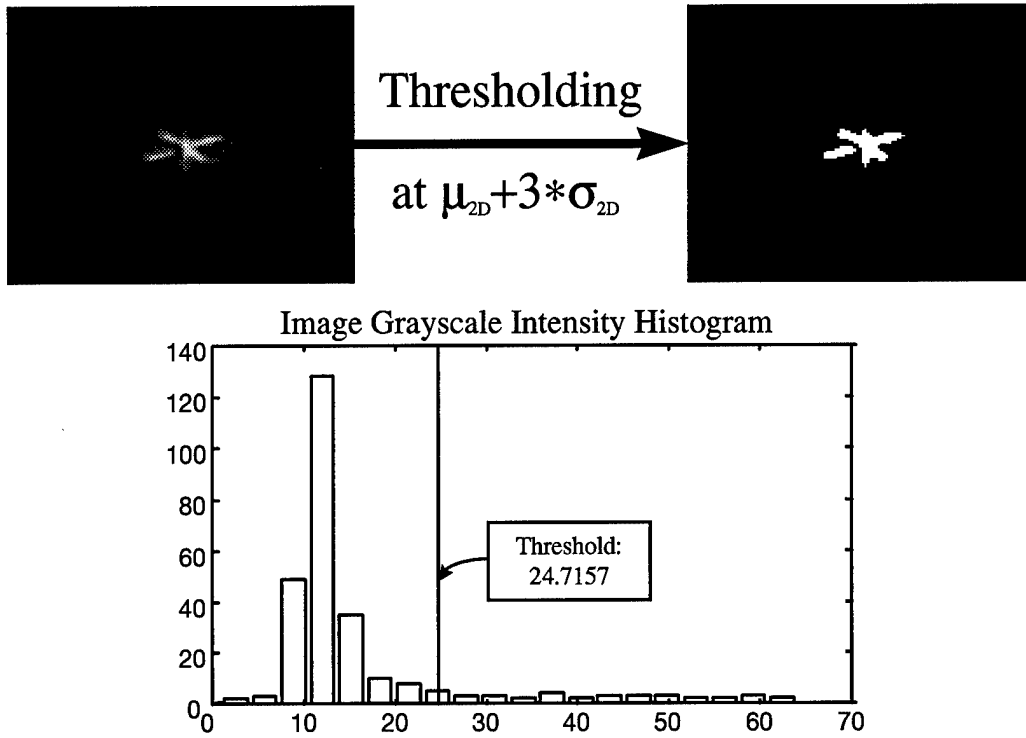
Figure 17. Real image gray scale to binary image transform by thresholding.

The simulated data required slightly different treatment than real data because the artificial nature of the background inflated the standard deviation. Thus, for the simulated images, a threshold value of mean plus one quarter of the deviation is used.

3.3.2.2 Moments of a Discrete Binary Image. By simplification of the gray scale definition of moments, we can achieve the following equation for the moments of a discrete binary image [11]:

$$m_{p,q} = \sum_{S} (x - \bar{x})^p (y - \bar{y})^q ,$$

where $(\bar{x}, \bar{y})$ denote the indices of the object center of mass, and S is the set of all pixels of value 1. From the definition of binary image moments, we can see that the zero-order moment $(m_{0,0})$ is simply a summation of the pixels, with the result being object area. The first-order moments $(m_{0,1}, m_{1,0})$ are defined to equal zero. Second-order moments $(m_{1,1}, m_{2,0}, m_{0,2})$ can be used to determine the axis around which the object can be rotated with minimum inertia with the equation [11,12]:

31

$$\varphi = \frac{1}{2}\arctan\left(\frac{2m_{1,1}}{m_{2,0} - m_{0,2}}\right)$$

The angle φ is the orientation of the axis of minimum inertia with respect to the x axis. If the object were an ellipse, the axis of minimum inertia would correspond to the major axis.

A measure of the circularity of the object is defined as eccentricity [11,12]:

$$\varepsilon = \frac{(m_{2,0} - m_{0,2})^2 + 4m_{1,1}{}^2}{(m_{2,0} + m_{0,2})^2}$$

Eccentricity ranges from zero for a circular object to one for a linear object, and is one of the five features resulting from the second-order moment analysis. The five measures of shape that can be included in the feature vector are eccentricity, angle of the axis of minimum inertia, and the three normalized second-order moments ($m_{1,1}$, $m_{2,0}$, $m_{0,2}$) [11,12]. Figure 18 depicts an example of thresholding and moment-based analysis of two, obviously normal and anomalous, simulated image pairs.
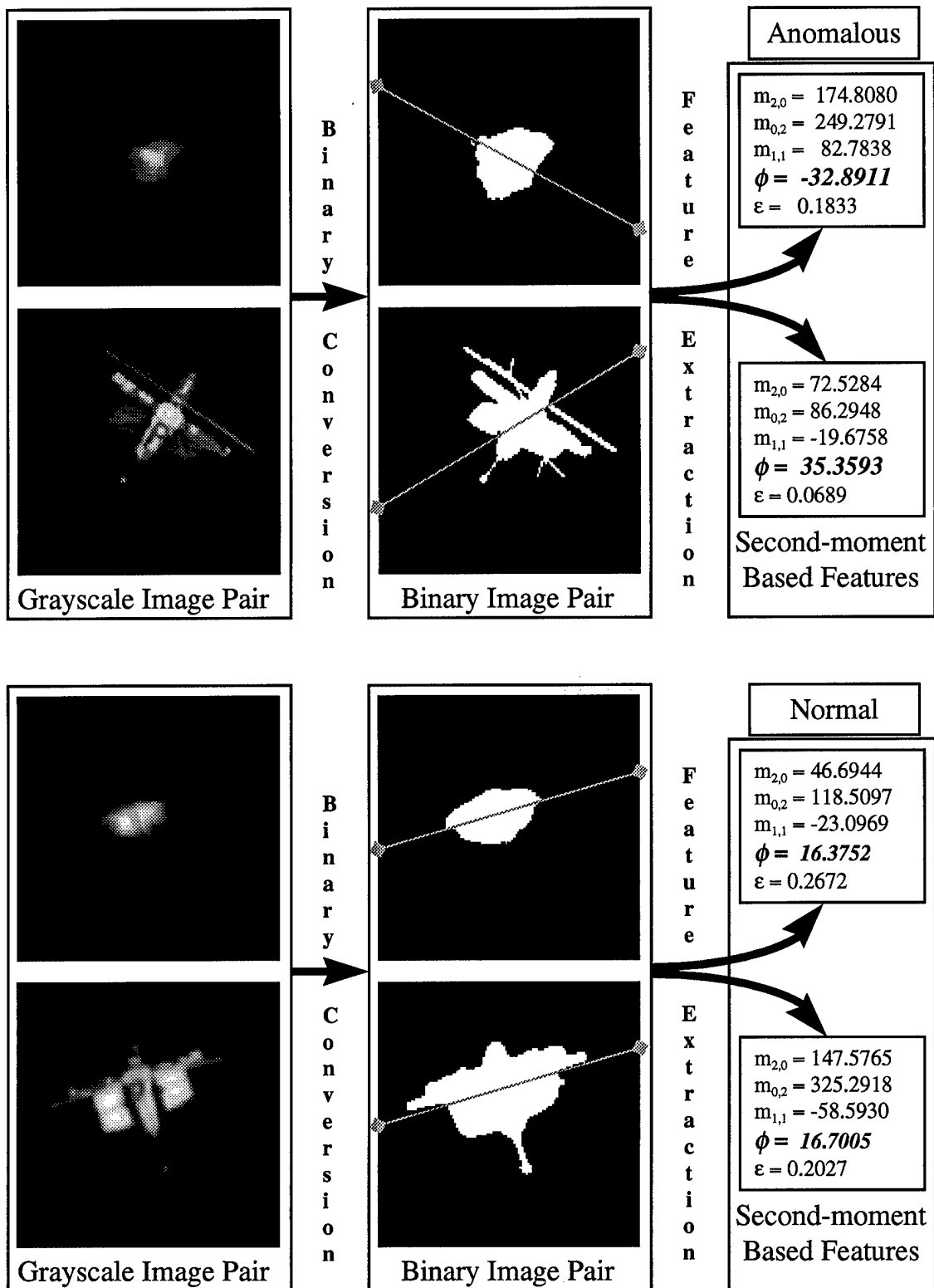
Figure 18. Moment-based analysis of normal and anomalous image pairs. The gray line shows the axis of minimum rotational inertia.

Moment-based analysis results in two feature sets, one for each of the images in the image pair.

3.3.3 Comparison Based Features.

The nature of this problem suggests features derived from a comparison of the images within an image pair. Sixteen feature sets are created with features extracted from comparisons within image pairs. Three Euclidean distance feature sets, each containing one feature, measure the distance between the block, wedge, and moment-based feature sets for each image pair:

$$\|M,R\| = \sqrt{\sum (M_i - R_i)^2} \quad \text{for } i = 1... \text{ length of feature set,}$$

where M is the feature set from the measured image and R is the feature set from the corresponding reference image.

A single feature set consisting of one value is created by taking the absolute difference between the angle of minimum inertia for the measured and reference images. This feature would certainly be effective for the example of Figure 18.

The remainder of the feature sets are constructed using the covariance operator to produce a covariance matrix of the feature sets from each image in an image pair. The diagonal elements of the covariance matrix are extracted for three feature sets. The diagonal elements of the matrix correspond to the variance between the two images in an image pair for each feature in the feature set. Because two images are being compared, the diagonal elements of the covariance matrix can be calculated:

$$D_i = \left(\frac{Fm_i + Fr_i}{2} - Fm_i\right)^2 + \left(\frac{Fm_i + Fr_i}{2} - Fr_i\right)^2 \quad \text{for } i = 1... \text{ length of feature set,}$$

where Fm_i is feature i within the measured image feature set, Fr_i is feature i within the reference image feature set, and D_i is the corresponding diagonal element of the covariance matrix. The Fr_i and Fm_i come from the feature sets defined in Sections 3.3.1 and 3.3.2. [13]

The diagonal elements of the covariance matrix make up three feature sets, with the statistically normalized diagonal elements forming an additional three. The number of features in each of these feature sets depends upon the feature set upon which the

34

covariance matrix was based: 25 for the block features, 8 for the wedge features, and 5 for the moment-based features.

The final six feature sets consist of one feature each. These feature sets are simply the sum of the diagonal element feature sets. Table 2 depicts the feature sets and their makeup.

Table 2. Feature set description.

Feature Set	Description of Feature Set	Number of Elements
1	Diagonal Elements of Covariance Matrix of Block FFT Features	25
2	Diagonal Elements of Covariance Matrix of Moment-based Features	5
3	Normalized Diagonal Elements of Covariance Matrix of Moment-based Features	5
4	Normalized Diagonal Elements of Covariance Matrix of Block FFT Features	25
5	Diagonal Elements of Covariance Matrix of Wedge FFT Features	8
6	Normalized Diagonal Elements of Covariance Matrix of Wedge FFT Features	8
7	Absolute Difference between Angles of Minimum Inertia	1
8	Euclidean Distance between Block FFT Features	1
9	Euclidean Distance between Moment-based Features	1
10	Euclidean Distance between Wedge FFT Features	1
11	Measured Image Block FFT Features	25
12	Measured Image Moment-based Features	5
13	Measured Image Wedge FFT Features	8
14	Normalized Measured Image Block FFT Features	25
15	Normalized Measured Image Wedge FFT Features	8
16	Normalized Reference Image Block FFT Features	25
17	Normalized Reference Image Wedge Features	8
18	Reference Image Block FFT Features	25
19	Sum of Diagonal Elements of Covariance Matrix of Moment-based Features	1
20	Sum of Normalized Diagonal Elements of Covariance Matrix of Block FFT Features	1
21	Sum of Diagonal Elements of Covariance Matrix of Wedge FFT Features	1
22	Reference Image Moment-based Features	5
23	Reference Image Wedge FFT Features	8
24	Sum of Normalized Diagonal Elements of Covariance Matrix of Moment-based Features	1
25	Sum of Normalized Diagonal Elements of Covariance Matrix of Wedge FFT Features	1
26	Sum of Diagonal Elements of Covariance Matrix of Block FFT Features	1

3.3.4 Choosing Features. From Table 2 it is evident that a number of different combinations of feature sets with a wide range of final feature vector lengths are possible. As complete enumeration of all possible feature set combinations is computationally unfeasible, only combinations of two feature sets will be completely tested. The results from the two feature set combinations will be examined to determine what feature sets have a tendency to improve classification accuracy. These feature sets will then be tested in an *ad hoc* manner in groups of greater than two.

Once the feature sets have been tested on the first data set, and the number of possible feature types reduced, the second data set will be used to continue selecting the best features. An analysis of each possible feature will reduce the number of features in the final feature vector. The final feature vector will contain the features that achieve a balance of good classification accuracy, low classification variance resulting from varying training data, and reduced dimensionality.

3.4 Classification Technique

Classification is the point in the pattern recognition process where we are required to make decisions. In the solution procedure depicted in Figure 5, two decision points exist. First, we must decide if a given image pair is a match, then we must decide if the satellite is behaving normally or abnormally based on the images that make up that pass.

3.4.1 Image Pair Classification. A Gaussian classifier is generally the first technique applied to classification problems and is capable of defining quadratic boundaries between classes [1]. If the statistical classifier works well enough, there is no need to look at other techniques. For initial classification of the image pairs, the Gaussian classifier performs well. The Gaussian classifier used is based on a minimum Mahalanobis distance decision boundary.

Training and test sets for data set 1 are made up of 400 image pairs. The training and test sets each contain 200 image pairs, which correspond to 10 satellite passes. The passes are balanced between normal and anomalous passes. Multiple perturbations of the training and test sets are used to evaluate data dependency. With multiple repetitions, a

confidence interval is also constructed for the true proportions of proper classification and the variability from random partitioning of the training and test sets can be evaluated.

Using data sets 2 through 7, we examine the impact of training and testing with each of the 6 different distortion levels to get a feel for the robustness to distortion. In the robustness experiments, the data are evenly split between training and testing. Data sets 2 through 7 are also used to explore the quantity of data required for proper training. This experiment requires a split of the data into training and test sets of different sizes.

3.4.2 Satellite Behavior Classification. Classification of a satellite's behavior is based on the results of every image pair in that satellite's pass. In the simulated data, each pass consists of 20 image pairs. Operationally, the passes can contain anywhere from 3 to 60 images, so the classification technique must not be specific to 20 image pairs. As an initial, basic, method of analyzing a pass, a simple "majority wins" strategy will be applied. If the pass contains more image pairs that are deemed matches than those that are not, the satellite will be considered to be behaving properly.

3.5 Experiments

The purpose of the experiments is to validate the solution procedure proposed in Figure 5. To accomplish this, we must find a set of features that properly classify satellite passes. The most extensive set of experiments concentrate of finding these features. The features should provide good classification accuracy, and reduce the dimensionality of the data representation. They should also produce a classifier that has low variability to the order of presentation of the training data. The experiments start with a broad view and explore the types of features that work. The different types of features explored are represented by the different feature sets in Table 2. The feature set experiments are conducted with a random separation of the data into test and training sets of equal size. Image pairs from the same satellite pass can be in both test and training sets in these experiments.

After exploration of feature type, the individual features are evaluated. These experiments are performed with random separation of the data into training and test sets

by satellite pass. In these experiments, image pairs from a single satellite pass are wholly contained in either the test or the training group.

Two other experiments are conducted: an experiment to explore robustness to data quality is performed, and an experiment to explore the quantity of data that should be used for training is conducted. The robustness experiment splits the data by satellite pass, while the quantity experiment does not.

4. Results

Results from the experiments conducted are reported in this chapter.

4.1 Feature Experiments

Determining what features should be used starts with an investigation of different types of features by examining the feature sets. The different feature sets described in Table 2 are combined in groups of two and tested. For each combination of feature sets, the classification accuracy of that feature set combination is recorded in three specific values: P_{gg} (the probability that an image pair that matches is classified as a match), P_{bb} (the probability that an image pair that does not match is classified as not a match), and CA (the summed classification accuracy, $P_{gg} + P_{bb}$).

The confusion matrix is also an effective method for looking at classification accuracy (Figure 19).

True Class

		Match	Not a Match
	Match	P_{gg}	$P_{bg}=100 - P_{bb}$
Classification	Not a Match	$P_{gb}=100 - P_{gg}$	P_{bb}

Figure 19. Confusion Matrix Defined.

Before we start looking at confusion matrices for different feature vectors, a general idea of which feature sets work well is needed. By complete enumeration, it is possible to look at pair-wise combinations of all the feature sets. By graphing the outcomes of each combination, a feel for the which feature sets are better can be attained (Figure 20). High and low overall classification accuracy trends can be seen in the CA surface.
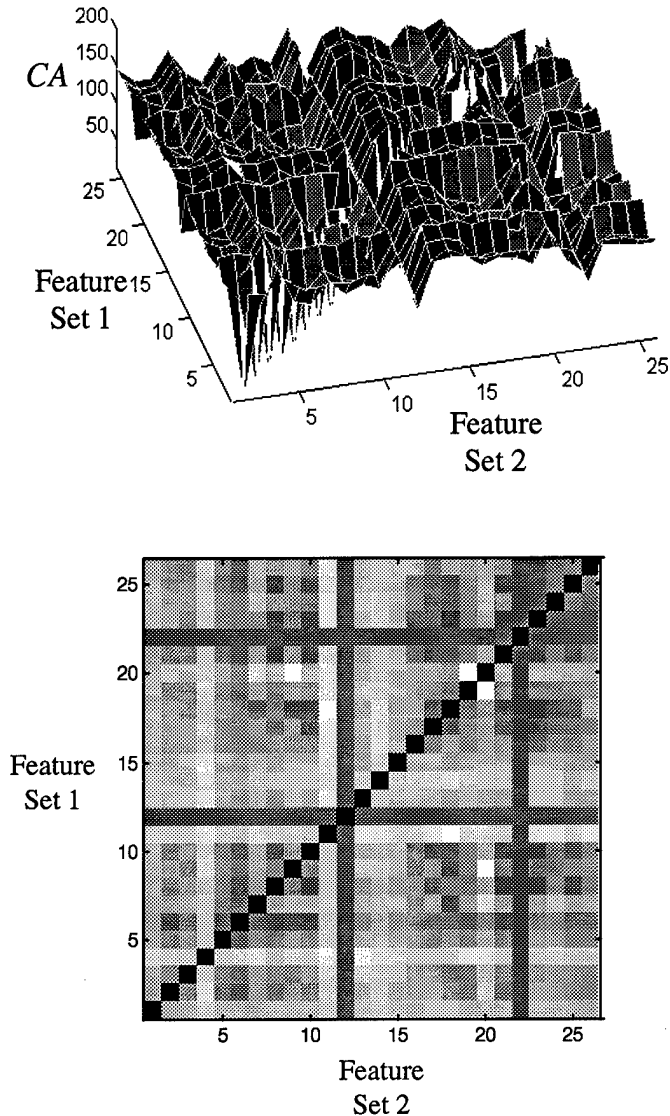
Figure 20. Summed classification accuracy (CA) surface and projection for a two feature set feature vector.

From the surface in Figure 20 we can see the relative performance of different feature sets in pair-wise combinations. The projection maps the surface onto a gray scale two-dimensional representation in which white represents the highest classification accuracy. With a general feel for which feature sets work better from the graphing heuristic (Figure 20), new combinations of multiple feature sets are created and tested.

Feature Sets 9, 20

True Class

	Match	Not a Match
Match	$\mu = 85.71$ $\sigma^2 = 3.8984$	$\mu = 11.84$
Not a Match	$\mu = 14.25$	$\mu = 88.16$ $\sigma^2 = 3.1066$

Feature Vector Length: 2

Feature Sets 11, 18

True Class

	Match	Not a Match
Match	$\mu = 73.19$ $\sigma^2 = 7.57$	$\mu = 0.01$
Not a Match	$\mu = 26.81$	$\mu = 99.99$ $\sigma^2 = 0.1$

Feature Vector Length: 50

Feature Sets 19, 20

True Class

	Match	Not a Match
Match	$\mu = 85.81$ $\sigma^2 = 3.8604$	$\mu = 11.56$
Not a Match	$\mu = 14.19$	$\mu = 88.44$ $\sigma^2 = 2.996$

Feature Vector Length: 2

Feature Sets 4, 14, 16, 20

True Class

	Match	Not a Match
Match	$\mu = 72.97$ $\sigma^2 = 8.4021$	$\mu = 0$
Not a Match	$\mu = 27.03$	$\mu = 100$ $\sigma^2 = 0$

Feature Vector Length: 76

Feature Sets 7, 19, 20

True Class

	Match	Not a Match
Match	$\mu = 84.23$ $\sigma^2 = 7.15$	$\mu = 8.13$
Not a Match	$\mu = 15.77$	$\mu = 91.87$ $\sigma^2 = 3.6918$

Feature Vector Length: 3

Feature Sets 4, 11, 18, 20

True Class

	Match	Not a Match
Match	$\mu = 76.1$ $\sigma^2 = 7.9347$	$\mu = 0$
Not a Match	$\mu = 23.9$	$\mu = 100$ $\sigma^2 = 0$

Feature Vector Length: 76

Figure 21. Selected confusion matrices for image pair classification with different feature sets as defined in Table 2.

42

Feature Sets 7, 8, 19, 20

Classification \ True Class	Match	Not a Match
Match	$\mu=74.45$, $\sigma^2=13.8669$	$\mu=6.65$
Not a Match	$\mu=25.55$	$\mu=93.35$, $\sigma^2=3.3739$

Feature Vector Length: 4

Feature Sets 13, 23, 19, 20

Classification \ True Class	Match	Not a Match
Match	$\mu=67.05$, $\sigma^2=8.4296$	$\mu=.61$
Not a Match	$\mu=32.95$	$\mu=99.39$, $\sigma^2=.7092$

Feature Vector Length: 18

Feature Sets 8, 9, 10

Classification \ True Class	Match	Not a Match
Match	$\mu=20.85$, $\sigma^2=7.81$	$\mu=3.21$
Not a Match	$\mu=79.15$	$\mu=96.79$, $\sigma^2=2.9035$

Feature Vector Length: 3

Feature Sets 4

Classification \ True Class	Match	Not a Match
Match	$\mu=70.84$, $\sigma^2=6.45$	$\mu=8.88$
Not a Match	$\mu=29.16$	$\mu=91.12$, $\sigma^2=2.9756$

Feature Vector Length: 25

Feature Sets 7, 20

Classification \ True Class	Match	Not a Match
Match	$\mu=85.39$, $\sigma^2=5.9813$	$\mu=13.47$
Not a Match	$\mu=14.61$	$\mu=86.53$, $\sigma^2=4.3681$

Feature Vector Length: 2

Feature Sets 11, 13, 18, 23

Classification \ True Class	Match	Not a Match
Match	$\mu=70.58$, $\sigma^2=7.8819$	$\mu=0$
Not a Match	$\mu=29.42$	$\mu=100$, $\sigma^2=0$

Feature Vector Length: 66

Figure 21. Selected confusion matrices for image pair classification with different feature sets as defined in Table 2., continued.

43

The means and variances in the confusion matrices in Figure 21 are based on 100 classification runs with random partitioning of the test and training sets. The standard deviation is greater if the classification results have more data dependency. Data dependency is something we hope to avoid with a good feature set, so high variances are not desirable.

Along with a desire for smaller variance, we would like to see good classification accuracy, and a small number of features. We are capable of achieving good classification accuracy with a very small number of features when using summed combination based feature sets. The combination features based on the block and moment based features performed well, and are thus explored in greater detail.

To continue exploration of the features, the remaining data sets, each based on the same 2000 images, are used. An expanded search into the block and moment-based comparison features is conducted as well as an evaluation of performance of the features over ranges of image distortion. With the six different sets of data at different distortion levels, a measure of robustness to distortion can be obtained.

The block and moment-based comparison features presented in Table 3 are considered for use in the final feature vector. Analysis of these features is initially accomplished heuristically by observation of the feature's distributions. Tests are performed on those features that appear to be good class differentiators.

Table 3. Feature descriptions.

Features	Description of Features
1-25	Magnitude of the Difference between the Thresholded Pseudo-Measured and Reference images Block FFT features.
26-30	Magnitude of the Difference between the Pseudo-Measured and Reference images Moment-Based features.
31	Euclidean distance between the Thresholded Pseudo-Measured and Reference images Block FFT features.
32	Euclidean distance between the Pseudo-Measured and Reference images Moment-Based features.
33-57	Normalized Diagonal Elements of Covariance Matrix of Thresholded Pseudo-Measured and Reference images Block FFT features.
58-62	Normalized Diagonal Elements of Covariance Matrix of Pseudo-Measured and Reference images Moment-Based features.
63	Summed absolute value of the Normalized Diagonal Elements of Covariance Matrix of Thresholded Pseudo-Measured and Reference images Block FFT features.
64	Summed absolute value of the Normalized Diagonal Elements of Covariance Matrix of Pseudo-Measured and Reference images Moment-Based features.
65	Euclidean distance between the Magnitude of the Thresholded Pseudo-Measured and Reference images Block FFT features.

To find good features for a statistical classifier, an examination of the training data's statistical distributions is a good heuristic with which to start. The following histograms are built with features extracted from the 500 matching image pairs and 500 non-matching image pairs in the training data (Figure 23). The superimposed normal probability curves show the Gaussian representation of the data used in the statistical classifier. There are three histograms for each feature examined. The first two show the distribution of each class (matching / non-matching) with the corresponding Gaussian representation. The third histogram for each feature shows the intersection of the two features' Gaussian representation, and give a feel for the overlap of the classes in that feature.
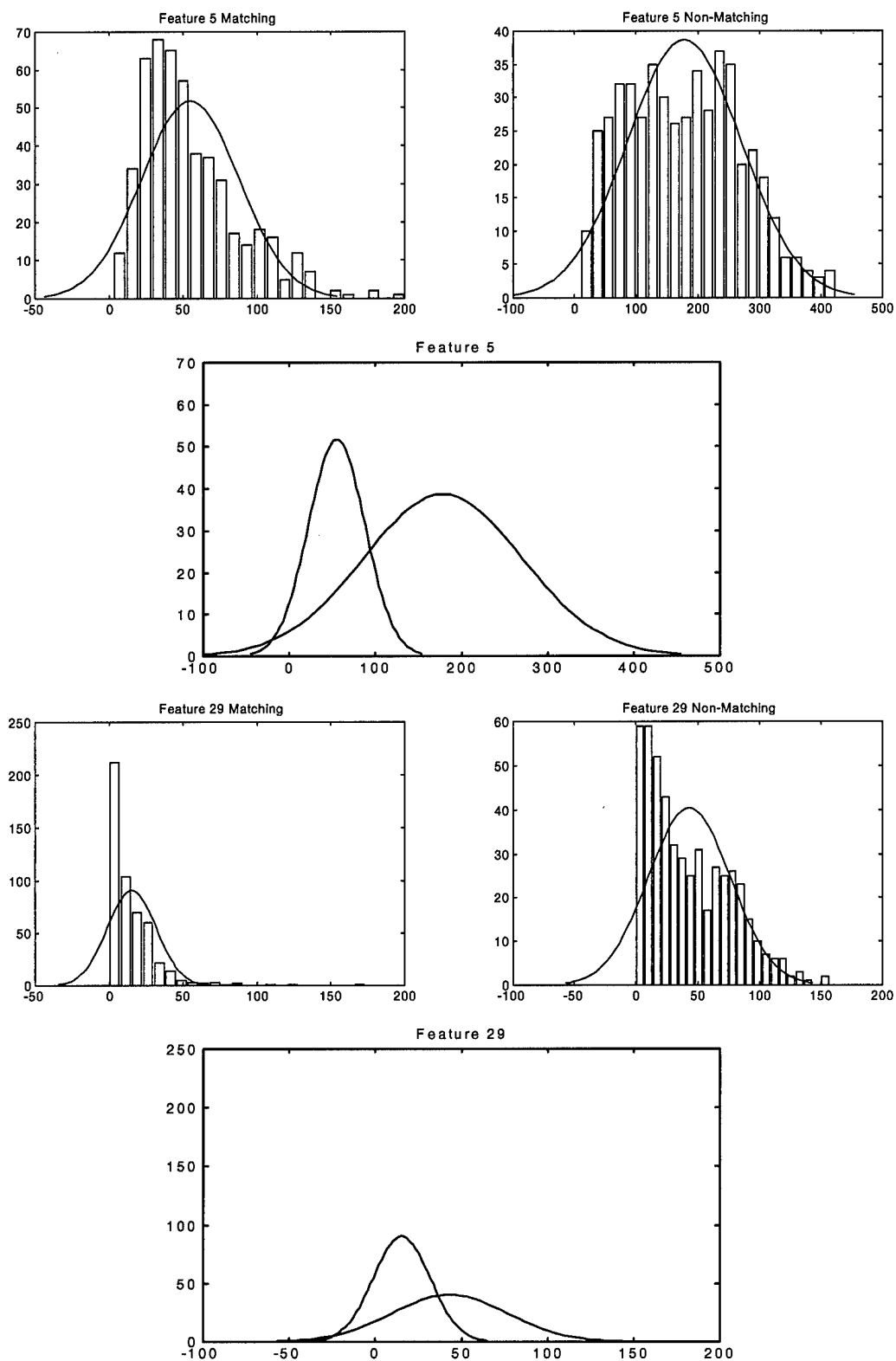
Figure 22. Selected feature histograms.
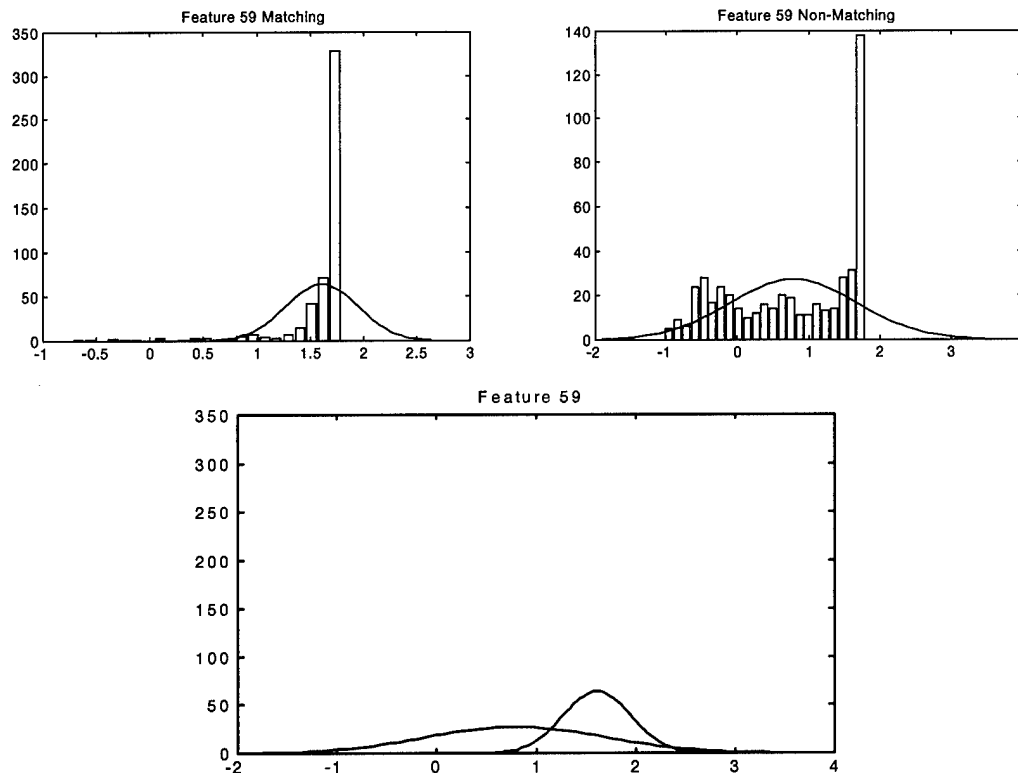
Figure 23. Selected feature histograms, continued.

An evaluation of the histograms of all 65 possible features identifies features 5, 29, and 59 as having good class separation. While these three features do not provide the best classification accuracy achieved, the features make intuitive sense and should work well with real data. The accuracy for image pair classification with these three features is high enough that satellite pass classification accuracy is very high. The confusion matrix in Figure 24 is based on images distorted with the random OTF analyzed with 1000 variations of data partitioning. The data are partitioned by satellite pass.

	True Class	
Features 5, 29, 54	Match	Not a Match
Classification — Match	$\mu = 73.941$ $\sigma^2 = 4.6662$	$\mu = 14.558$
Classification — Not a Match	$\mu = 26.059$	$\mu = 85.442$ $\sigma^2 = 3.3694$

Feature Vector Length: 3

Figure 24. Confusion matrix for features 5, 29, and 59 with 1000 perturbations of the data.

Feature 5 is the energy difference between one block feature extracted from the measured and reference images. Feature 29 is the difference in the angle of minimum inertia between the measured and reference images. Feature 54 is the diagonal element of the covariance matrix of the measured and reference images that corresponds to one of the second-order moments. With these three features, the dimensionality of the problem is dramatically reduced, and the variability of classification accuracy due to data division is small. These three features achieve the stated objectives, and will be used as the final feature vector in evaluating robustness to distortion and training set size.

4.2 Robustness Experiment

To evaluate the solution technique's robustness to atmospheric distortion, data sets 2 through 7 are used. With these data sets, training and testing is accomplished at each OTF level for a total of 36 test points. The experiment is performed 1000 times to quantify the training data dependency. For each of the 1000 repetitions, the 100 passes are randomly divided into training and test sets. The training and test sets each contain 50 passes: 25 anomalous and 25 normal. Figure 25 depicts the results from this 36 point test

in a gray scale representation. Better classification accuracy is represented by lighter
squares. From the overall shade of a column, we can see how well training with a
particular OTF works.
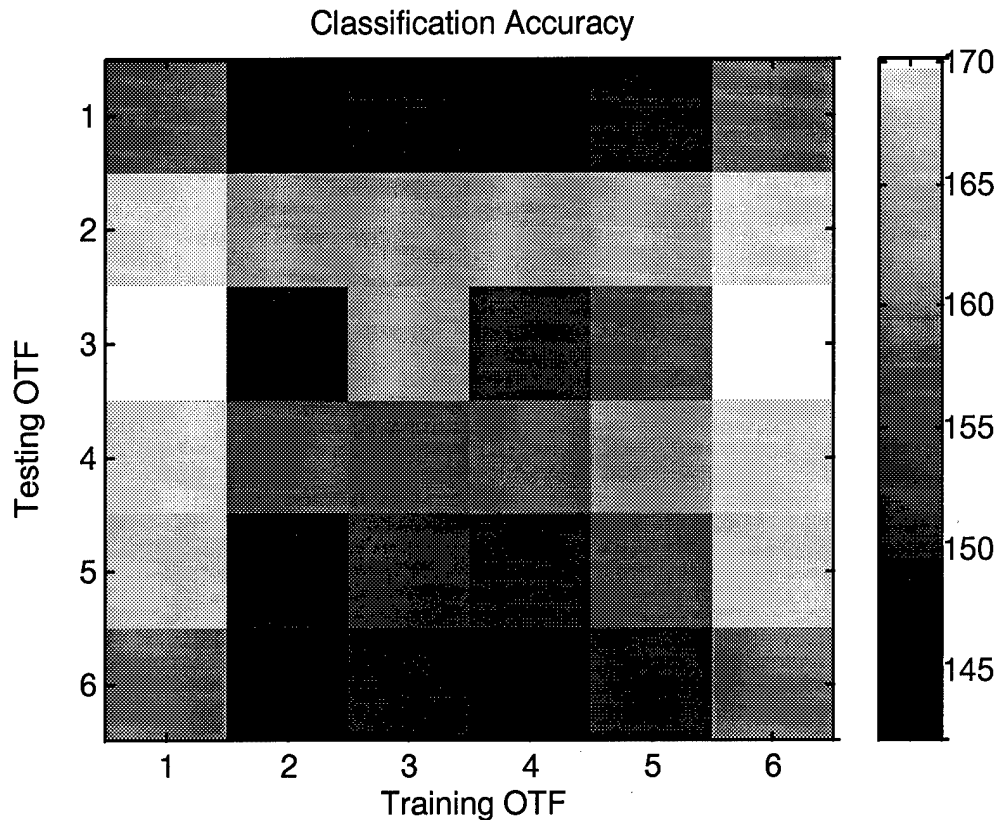
Figure 25. OTF level test results.

As we can see from the general light shade in columns 1 and 6, the best
performance is achieved when training is accomplished with data created using OTF 1
(low distortion) or OTF 6 (random distortion) (Figure 25). Low or varying distortion in
the training set results in good accuracy in classification image pairs distorted at any
level. Classification is actually better for images distorted to a different level than for
images at the training distortion level. It appears from the results of training with OTF
level 1, that training with more accurate data results in better generalization, which allows
better classification regardless of the distortion. Evaluation of the random OTF levels
indicates training with a varying level also provides a more generalized classifier.

So, when trained with a level of distortion that allows good generalization, the classifier is robust to distortion. Training with real data should lead to a generalized classifier as the real image pairs would have a high degree of randomness in the level of distortion. A classifier trained with real data would be similar to the classifier trained with the random OTF.

4.3 Data Quantity Experiment

Exploring the data quantity requirement is the final experiment conducted. Using the data from the random OTF, nine different partitions of the data into training and test sets are made. Classification accuracy is recorded at each level. The data are split into sets: 10% training / 90% test, 20% training / 80% test, ..., 90% training / 10% test. 100 repetitions with random assignment to test and training groups are accomplished. Figure 26 contains a plot of the results.
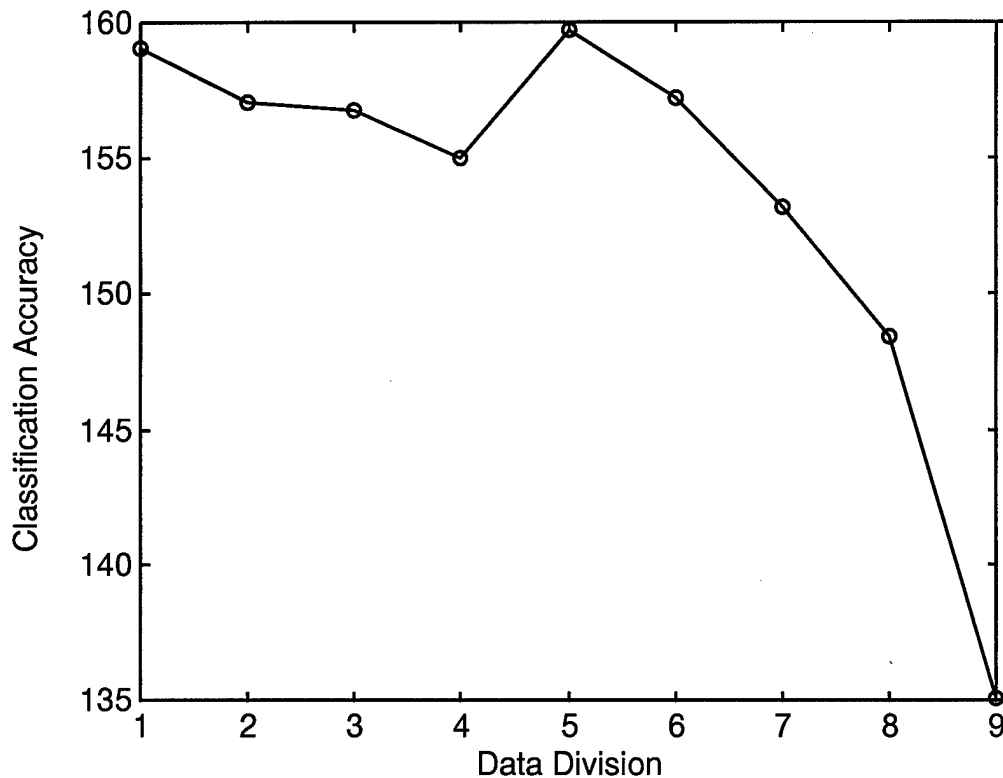

Figure 26. Classification accuracy as affected by data split. Data division * 10 = percentage used for training.

These results indicate that the classifier can be over-trained. It appears that a smaller amount of training data results in a more generalized classifier. The main loss of classification accuracy was due to a loss of P_{gg}, or proper classification of matching pairs. As more training exemplars are used, the ability of the classifier to properly identify a good pass lessens. While the specific number of training exemplars for real data may be quite different than for the manufactured data, care must be taken not to over-train the classifier when using real data. A similar test can be performed with real data to establish the amount of training data that does not reduce the generality of the classifier.

4.4 Classification Accuracy for Image Pairs

Using the generalized classifier based on training and testing with the random OTF, and a feature vector composed of the three features from above, we can properly identify image pair matches with 74% accuracy. Classification accuracy for non-matching pairs is better: 85.5%.

With 1000 random partitions of the data, and an, apparently valid (Figure 27), assumption of normality, we can calculate confidence intervals for the containing true probabilities of classification to be:

$$P_{gg} = 73.941 \pm .1366,$$
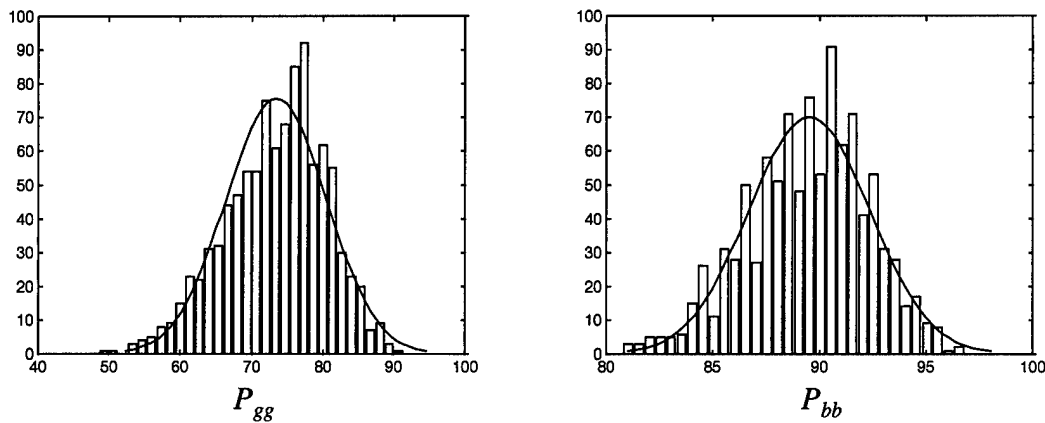
$$P_{bb} = 85.442 \pm .1161.$$

Figure 27. Histogram of 1000 classification results with superimposed normal curve.

4.5 Classification Accuracy for Satellite Behavior

Image pair classification is the building block for the true endeavor: classifying a satellite's behavior as "normal" or "anomalous" over a single pass. This can be accomplished with a simple maximum count algorithm that can be analyzed as a binomial trial experiment. If 50% or more of the image pairs in a pass match, we will assign the satellite behavior to the "normal" class, otherwise the satellite behavior will be flagged as "anomalous." This experiment, while simple, yields very good results.

The most egregious error that could be made would be to call a satellite behaving abnormally, normal. Analyzing this situation with as a binomial trial, we can define the probability of success as being the probability that an image pair that does not match is classified as matching. Using the equation for a binomial experiment and an assumption of 20 image pairs per pass, it is simple to calculate the probability of misclassifying an anomalous satellite [14]:

$$p[\text{normal} \mid \text{anomalous}] = \sum_{y=11}^{20} \binom{20}{y} P_{bg}{}^{y} P_{bb}{}^{20-y}$$

where $P_{bg} = 1 - P_{bb}$, and $P_{bb} = .85442$.

$p[\text{normal} \mid \text{anomalous}] = .000029024$

So, with 20 images, there is virtually 100% chance of correctly classifying an anomalously behaving satellite. Similarly, the chance of correctly classifying a normally behaving satellite is 99.5%. With only 2 image pairs, the maximum pick algorithm gives a probability of misclassifying an anomalously behaving satellite of only 0.0212. Figure 28 shows the high power of the simple maximum pick algorithm based on the number of image pairs in a pass.
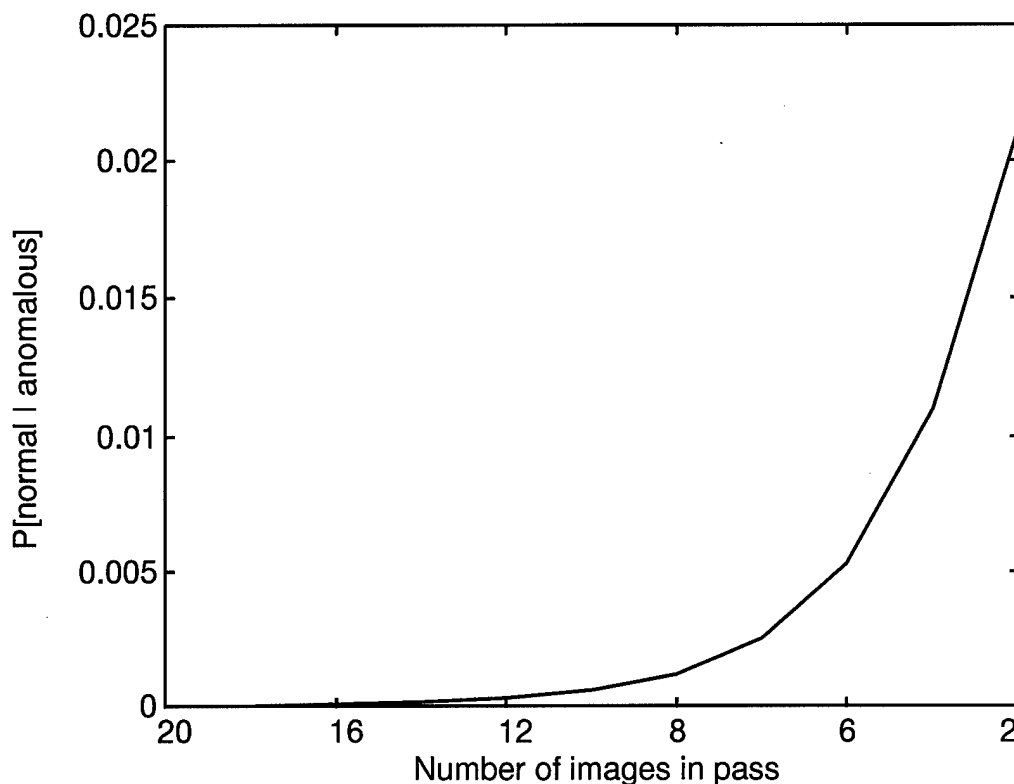
Figure 28. Probability of misclassification of an anomalous satellite pass plotted against the number of images in the pass.

The probabilities calculated above assume independence of classification for each image pair. From experiments with the simulated data, this assumption does not appear to be valid. The contrived nature of the data gives rise to data in which the conditions leading to misclassification of one image pair in a pass are present in other image pairs in that pass. If real training data are carefully chosen, this situation should be avoidable. This will lead to greater independence in classification of each image pair. It is also simple to change the counting threshold for finding a pass "normal" to a point that ensures 100% detection of anomalous passes.

5. Conclusions

This chapter makes high level conclusions regarding the solution procedure and discusses the operational application of this work.

5.1 Validity of Solution Procedure

The experimental results clearly show that the solution procedure proposed in this thesis works. By utilizing the information that is available with each measured image, it is possible to create a simulated reference image against which to compare the measured image. This reference image creation and comparison can be made in real time to ensure that the status of Air Force monitored satellites is almost constantly available.

While the specific results and specific conclusions that could be drawn from them are interesting, we must keep in mind that the experiments upon which the results are based use simulated data. The training sets are more limited in size than an operational system would use, but the simulated data are probably more consistent than real data would be. These results show only how well these algorithms work on simulated data, and while care has been taken to make the simulated data as real as possible, there is no escaping the fact that it is simulated data.

The specific results are important in comparing different methodologies, and understanding how well the solution methodologies would compare in an operational setting. As with any use of simulation, the relative outcomes are generally more telling than the specific values.

5.2 Operational Application

This thesis demonstrates that an algorithm using simulated reference imagery is feasible, but to apply this procedure as an operational system, real data must be collected for training purposes. Almost every step of the procedure would require modification to ensure performance was maximized for real data. Improvements could be made to the procedure in many ways.

A filtering system could be applied to remove images of poor quality before sending them to the classification algorithm. Quality could be defined as containing shape

information. Removing images of poor quality would reduce the number of images for the final maximum pick algorithm, but the images that remained would produce more accurate results.

Different classification techniques could be applied to the image pair classification problem. Many techniques exist that are not based on statistics, but even improved statistical techniques could be used. The probability distribution of some of the features is not best modeled by a Gaussian probability density function. Features based on differences, which have an exponential distribution could be modeled with the exponential distribution.

The technique used to classify satellite passes could be made more complex. The confidence in each of the image pair classifications could be used as a part of the pass classification scheme. This could give more weight to image pairs in a pass that really do or do not match. The image pairs that are close calls would not carry as much weight, which could make the classification of passes more accurate.

5.3 Final Summary

A more complete understanding of the problem motivated a new approach to take advantage of all available information. With more information, the problem is not as daunting and a simple solution procedure based on comparisons is possible. This procedure works on the simulated data, and proves that the concept will work if applied to the operational situation.

Bibliography

1. Bishop, Christopher M. *Neural Networks for Pattern Recognition.* New York: Oxford University Press, 1995.

2. Duda, Richard O. and Peter E. Hart. *Pattern Classification and Scene Analysis.* New York: John Wiley and Sons, 1973.

3. Fielding, Kenneth H. and Dennis W. Ruck. "Spatio-Temporal Pattern Recognition using Hidden Markov Models," *IEEE Transactions on Aerospace and Electronic Systems,* 31:1292-1300 (Oct 1995).

4. Rabiner, Lawrence R. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE,* 77(2):257-286 (1989).

5. Neiberg, Leonard and David P. Casasent. "Feature Space Trajectory (FST) Classifier Neural Network," *Proc. Soc. Photo-Opt. Instrum. Eng.* (SPIE), 2353:276-292 (1994).

6. Neiberg, Leonard and David P. Casasent. "Features Space Trajectory Neural Net Classifier," *Proc. Soc. Photo-Opt. Instrum. Eng.* (SPIE), 2492:361-372 (1995).

7. Bruegger, Neal W. *Space Object Identification Using Features Space Trajectory Neural Networks.* Masters Thesis, Air Force Institute of Technology, March 1997.

8. Bringham, Oran E. *The Fast Fourier Transform and its Applications.* New Jersey: Prentice-Hall, Inc. 1988.

9. Schulze, Kathy. *RDCSIM Reference Guide* Version 1.0. 1996.

10. Brandstrom, Gary W. *Space Object Identification Using Spatio-Temporal Pattern Recognition.* Masters Thesis, Air Force Institute of Technology, December 1995.

11. Jahne, Bernd. *Digital Image Processing Concepts, Algorithms, and Scientific Applications.* New York: Springer-Verlag. 1995.

12. Warfel, Matthew. Moments Calculation Macro. http://www.cee.cornell.edu/~mdw/.

13. Neter, John, and others. *Applied Linear Regression Models.* Irwin, Inc. 1996.

14. Wackerly, Dennis D. and others. *Mathematical Statistics with Applications.* Wadsworth Publishing Company. 1996.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE 12 March 1998	3. REPORT TYPE AND DATES COVERED Masters Thesis	
4. TITLE AND SUBTITLE Real Time Detection of Anomalous Satellilte Behavior from Ground-Based Telescope Images			**5. FUNDING NUMBERS**
6. AUTHOR(S) Geoffrey S. Maron			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology, WPAFB OH 45433-6583			**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT/GOA/ENG/98M-01
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Capt Bruce Stribling PL/OL-YY 535 Lipoa Pkwy, Suite 200 Kihei, Maui, HI 96753			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION AVAILABILITY STATEMENT Distribution Unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT *(Maximum 200 words)*

Air Force analysts are faced with the task of monitoring satellites with ground-based telescopes. Images are collected and analyzed in a time-consuming and subjective effort to detect any behavior that is anomalous. This research maximizes use of a priori information to create an automated, real-time satellite behavior classification tool.

Using modeling software and knowledge of a satellite's orbit, reference imagery is created for each measured image in a satellite pass. Features are extracted from the measured and reference image pairs that provide good overall gaussian classification accuracy (85%), reduce the dimensionality of the problem (from 32,768 down to 3), and are least dependent on data partitioning. The statistical image pair classifier is tested for robustness to atmospheric distortion, and training data requirements are explored.

Satellite behavior is classified by counting the classification results for the image pairs in a satellite pass. A binomial analysis of the classification technique predicts virtually 100% classification accuracy of satellite behavior. This research demonstrates the validity of model based satellite behavior analysis.

14. SUBJECT TERMS pattern recognition, gaussian classification, satellite imagery, feature extraction			15. NUMBER OF PAGES 67
			16. PRICE CODE

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