Pose Independent Target Recognition System

Using Pulsed Ladar Imagery

by

Alexandru N. Vasile

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

February 4, 2004

[June 2004]

Author____
Department of Electrical Engineering and Computer Science
February 4, 2004

Certified by_________ ____________________
Dr. Richard M. Marino
Company Thesis Supervisor

Certified by____________________ ______
M. Wells
ıpervisor

Accepted by____________________ ______
C. Smith
Chairman, Department Committee on Graduate Theses

This work is sponsored by the Defense Advanced Research Projects Agency under AF Contract #F19628-00-C-0002. Opinions, interpretations and conclusions are those of the authors and are not necessarily endorsed by the United States Government

Abstract

Although a number of object recognition techniques have been developed to process LADAR scanned terrain scenes, these techniques have had limited success in target discrimination in part due to low-resolution data and limits in available computation power. We present a pose-independent Automatic Target Detection and Recognition System that uses data from an airborne 3D imaging Ladar sensor. The Automatic Target Recognition system uses geometric shape and size signatures from target models to detect and recognize targets under heavy canopy and camouflage cover in extended terrain scenes.

A method for data integration was developed to register multiple scene views to obtain a more complete 3D surface signature of a target. Automatic target detection was performed using the general approach of "3D cueing," which determines and ranks regions of interest within a large-scale scene based on the likelihood that they contain the respective target. Each region of interest is then passed to an ATR algorithm to accurately identify the target from among a library of target models. Automatic target recognition was performed using spin-image surface matching, a pose-independent algorithm that determines correspondences between a scene and a target of interest. Given a region of interest within a large-scale scene, the ATR algorithm either identifies the target from among a library of 10 target models or reports a "none of the above" outcome.

The system performance was demonstrated on five measured scenes with targets both out in the open and under heavy canopy cover, where the target occupied between 1 to 10% of the scene by volume. The ATR section of the system was successfully demonstrated for twelve measured data scenes with targets both out in the open and under heavy canopy and camouflage cover. Correct target identification was also demonstrated for targets with multiple movable parts that are in arbitrary orientations. The system achieved a high recognition rate (over 99%) along with a low false alarm rate (less than 0.01%)

The contributions of this thesis research are: 1) I implemented a novel technique for reconstructing multiple-view 3D Ladar scenes. 2) I demonstrated that spin-image-based detection and recognition is feasible for terrain data collected in the field with a sensor that may be used in a tactical situation and 3) I demonstrated recognition of articulated objects, with multiple movable parts.

Immediate benefits of the presented work will be to the area of Automatic Target Recognition of military ground vehicles, where the vehicles of interest may include articulated components with variable position relative to the body, and come in many possible configurations. Other application areas include human detection and recognition for Homeland Security, and registration of large or extended terrain scenes.

Key Words: ATR, registration, detection, recognition, surface matching, 3D Ladar

Acknowledgements

I would like to thank my thesis advisors, Professor William (Sandy) Wells and Dr. Richard Marino for their advice and guidance throughout the duration of this research. Their constant support kept me focused and committed to accomplish my thesis research goals. I would also like to thank my Lincoln Lab group leader Dr. Richard Heinrichs for funding and enthusiastically supporting my research. Also, I would like to thank Professor Berthold K.P. Horn for his critique, suggestions, and advice given for Chapters 1, 2, 4 and 5.

Many of my colleagues helped me obtain my Master's degree. My officemate, Michael O'Brien, was a great source on the hardware aspects of the 3D Ladar sensor and the simulation of the Ladar data from target CAD models. I also found the help from Luke Skelly and Joe Adams invaluable on my numerous questions in C++. Also, Justin Libby helped me clarify my thoughts and ideas through our numerous chalk-talk and white-board discussions.

Finally, I would like to thank my family. My parents, Stefan and Eliza Vasile have been a source of constant support, encouragement and inspiration. On numerous occasions, they provided me with advice on how to conduct my thesis research based on their own research experience.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Massachusetts Institute of Technology

Department of Electrical Engineering and Computer Science

Proposal for Thesis Research in Partial Fulfillment

of the Requirements for the Degree of

Master of Engineering

Title: Pose Independent Target Recognition System Using Pulsed Ladar Imagery

Submitted by: Alexandru N. Vasile Signature of Author

Date of Submission: August 16, 2002
Expected Date of Completion: January 2004
Laboratory where Thesis will be done: MIT Lincoln Laboratory

Brief Statement of the Problem:

Although a number of object recognition techniques have been developed to process LADAR scanned terrain scenes, these techniques have had limited success in target discrimination in part due to low-resolution data and limits in available computation power [1]. Significant improvements in the resolution and accuracy of LADAR data as well as increases in computational power have opened the possibility to use more sophisticated techniques that can better discriminate targets. Several novel techniques have been implemented, but have only been successfully demonstrated on a limited class of scenes containing objects with relatively simple geometries [2].

The purpose of this thesis is to apply improved techniques to accurately recognize partially occluded, highly articulated targets, that are present in large data sets with high levels of clutter. An Automatic Target Recognition system will be built, which will include a target detection and target recognition stage. Two detection methods will be implemented, namely a geometric feature extraction method and a 3D-Cueing method [16] [3]. Target recognition will be implemented based on an improved spin-image method [24]. The fidelity of the results will be determined based on field collected and simulated LADAR data.

Supervisor Agreement:

The program outlined in this proposal is adequate for a Master's Thesis. The supplies and facilities required are available, and I am willing to supervise the research and evaluate the thesis report.

Professor William Wells Associate Professor, Harvard Medical School & Brigham and Women's Hospital, Member of the Faculty of the Harvard-MIT HST Division, MIT Thesis Supervisor	Dr. Richard Marino Staff, Lincoln Laboratory MIT-LL Supervisor

Chapter 1 Introduction

Three-dimensional Laser radar (3-D Ladar) sensors produce range images that provide explicit 3-D information about a scene. MIT Lincoln laboratory has actively developed the laser and detector technologies that make it possible to build three-dimensional Ladar that can capture an entire 3-D image on single pulse at a few centimeter range resolution [4]. The Lasers and Sensors Group at Lincoln Lab has built a functional 3-D Ladar system with a 32x32 array of APDs operating in Geiger mode. Recent field tests using the sensor have produced high-quality 3-D imagery of targets behind obscurants for extremely low signal levels [4].

The target detection and recognition algorithms implemented in this thesis use field data collected by this high-range-resolution sensor. The primary goal was to accurately detect and recognize targets present in large terrain scenes where the target may occupy less than 1% of the scene and have more than 200 points on target. A secondary system goal was to demonstrate correct target identification with foliage occlusion greater than 70%. Another goal was to demonstrate correct identification of articulated targets, with multiple movable parts that are in arbitrary orientations. The above goals have to be met while achieving a high recognition rate (over 99%) along with a low false alarm rate (less than 0.01%). Furthermore, in order to provide a system that might have a significant practical value for automatic target recognition under battlefield conditions, the recognition runtime performance on a standard personal computer was constrained to be less than 10 minutes.

The problem of automatic target recognition in Ladar range imagery has been an active topic of research for a number of years [5]. Automatic Target Recognition (ATR) involves two main tasks: target detection and target recognition [6]. The purpose of target detection is to find regions of interest (ROI's) where a target may be located. By locating ROI's, one is able to filter out a large amount of background clutter from the terrain scene, making object recognition feasible for large data sets. The ROI's are then passed to a recognition algorithm that identifies the target [6].

Target detection methods attempt to determine the presence of a target in a large data set by quickly filtering large portions of the scene prior to submitting the data to the recognition algorithm. In the ATR field, detection methods that reduce a large data set to a few regions of interest are known as "cueing" algorithms [3]. The application of a cueing algorithm as a data-preprocessing step allows for vastly reduced target recognition time.

Target detection approaches can be classified as image-based and model-based [7]. The traditional image-based approach is based on template matching: the target is separated from its surrounding area by extracting a silhouette based on a target image template [8]. However, silhouette extraction algorithms did not reliably recover the true silhouette from real imagery, thus seriously degrading the robustness of target detection [8]. In general, the template approach suffered from the complexity in finding the silhouette in the image, as well as the complexity of creating the template database [7].

With significant improvement in LADAR sensor accuracy and resolution, and increased computational power, detailed 3D structural information may be obtained from the data and used by model-based approaches. Traditional model-based approaches rely on boundary segmentation and planar surface extraction to describe the scene. Target detection is then performed through the use of trained neural networks or genetic algorithms [8] [9] [10] [11] [12]. One recent cueing algorithm that is applicable to large Ladar data sets is known as 3-D Cueing, developed by Owen Carmichael and Martial Hebert [3].

Given a region of interest, the recognition algorithm attempts to classify the particular target based on a library of target models. The target models are used to represent a unique signature that is present in the target data set. There are numerous ways to encode the target models. For Ladar data, where the scene data consists of an unstructured point cloud, object representation schemes fall into two categories: surface-based 3D model representations and shape-based 2D model representations.

Surface-based 3D model schemes perform geometrical surface matching between a library of 3D surface models and a data scene. Traditional 3D geometrical feature-matching algorithms segment the target into simple geometric primitives, such as planes, and cylinders and record the relative spatial relationship of each geometric primitive in the target model [13] [14] [15]. The scene is then segmented in the same manner, and searched for a group of primitive objects located in a similar spatial structure as in the target model [16][17]. Recent methods have shown that planar patch segmentation is robust to noisy range data [18]. In addition, current 3D feature grouping schemes have been proven to work even when the target is partially occluded [19].

An alternate approach to 3D geometric feature matching is to reduce the three-dimensional recognition problem to a set of two-dimensional recognition problems, where the target signature is encoded by a shape-based 2D representation. The primary advantage of the shape-based recognition approach over 3D geometrical matching is that it can scale well to large data sets with high levels of clutter [3] [20]. In addition, the recognition algorithms can benefit from the tremendous amount of work done in the relatively mature field of 2D image analysis. A couple of recent algorithms that use shape-based representations are Shantaram et al's contour-based algorithm, Dorai et al.'s shape spectra algorithm, Yamany et al.'s surface signatures and Andrew Johnson's spin-image algorithm [21] [22] [23] [24].

The remainder of Chapter 1 will discuss the current available detection and recognition algorithms. The most promising approaches for processing Ladar terrain data were chosen as the basis of the ATR implementation. Immediate benefits of this work will be to the area of Automatic Target Recognition of military ground vehicles, where the vehicles of interest possess articulated components with variable position relative to the body, and come in many possible configurations [2]. Another area of application can be in the context of interpreting and analyzing articulated human motion.

1.1 Background

This section will provide a background on the recent algorithmic advances in the ATR field. The advantages and disadvantages of each approach will be discussed and a justification will be given for using the spin-image surface matching process.

We analyzed and evaluated several shape-based methods for target recognition, namely Contour Matching, Shape Spectra, Surface Signatures and Spin Images [21] [22] [23] [24]. Shantaram et al's contour algorithm matches the outer and inner edge contours of a target model to contours found in the scene range data [21]. The method is heavily dependent on accurately capturing edge information in the data scene. From their results, the authors acknowledge that a recognition system based on edge detection is not very robust and cannot distinguish between very similar objects [21]. Thus, this method is not the best match for processing our particular data set for several reasons: 1) our Ladar data contains large occlusions due to camouflage and canopy cover, resulting in broken, partial boundaries and 2) edge detection on some of our LADAR data sets may perform poorly due to measurement noise.

The second algorithm considered is Dorai et al.'s shape spectra technique. The shape spectra algorithm can recognize 3D free form surfaces by matching view-based representations of the scene and model targets. [22] For the limited class of scenes considered, the algorithm has proven to be successful. However, this particular algorithm does not scale well to large data scenes with high levels of clutter [3]. Since our LADAR data sets can include heavy clutter such as canopy cover in proximity of the target of interest, the shape spectra algorithm would not be an effective object recognition solution.

The next algorithm considered is Sameh Yamany et al.'s surface signature technique. The surface signatures algorithm creates a 2D signature image for each surface point. The signature image encodes the surface curvature as seen from this point. The task of surface-signature object recognition is to find target to model point correspondences based on the similarity of a target surface signature to a model surface signature [23]. The surface curvature is encoded by two parameters: 1) the distance from the signature point basis to another point in the data set and 2) the angle between the surface normal at the signature point basis and the vector connecting the point basis to another point in the data set. The 2D representation is thus heavily dependent on correct normal orientation and also somewhat dependent on the relative location of neighboring points. For instance, if there are any noisy points close to the point basis of the signature, the signature will have a wide distribution of vertex to normal angles, which will be mapped to very different locations in the given 2D space. The representation's sensitivity to noisy points and normals may degrade the matching of target signatures to model signatures. In addition, the author mentions that target scale issues have to be resolved at the recognition stage of the algorithm [23].

The fourth algorithm considered was Andrew Johnson et al.'s spin-image method. Similar to surface signatures, spin-images capture a 2D representation of the object from an individual point basis. The task of spin-image object recognition is to find good point correspondences between the scene data set and the model data set by finding similar target and model spin-images. Given an oriented point (a 3D point with a surface normal) the data set is reduced to the following 2D parameter space: 1) the perpendicular distance to the line through the surface normal and 2) the signed perpendicular distance to the tangent plane defined by the oriented point normal and position [24]. Similar to the surface signature approach, the spin-image representation is somewhat dependent on the point surface normal. However, an important advantage of the spin-image over the surface-signature approach is that small changes in the relative point location do not have a significant effect on the spin-image, since a noisy surface point measurement will be mapped close to the 2D spin-image coordinates of an ideal point that contains no noise. Therefore, the spin-image representation should be much more robust to measurement noise [24]. Furthermore, spin-images can easily address clutter problems. Since the 2D parameters of a spin-image are distance-based, spin-images can be analyzed on a local to global scale. [24] For highly cluttered data scenes, spin-images can be compared within a smaller distance to the point basis, reducing the spin-image to a local representation of the object [25]. Aside from the better 2D representation of spin-images as compared to surface signatures, a significant amount of work has been done in spin-image recognition that demonstrates the overall robustness of the spin-image representation over the previous methods [24] [25] [26] [27].

Two model-based algorithms were considered for target detection, namely a traditional geometric feature-based algorithm and Owen Carmichael et al's 3-D Cueing algorithm. The geometric feature based method uses planar extraction to separate possible targets from the background terrain. However, as mentioned in [20], the geometric feature-based method is not very robust to occlusion or clutter. Thus, the algorithm is not well suited for recognition of targets that are surrounded by a relatively high amount ground clutter and are underneath heavy canopy and camouflage cover.

The alternate approach is 3-D Cueing, which compares model signatures to scene signatures through a classifier that assigns weights to points, based on their likelihood that they are part of the target model. Points with a probability lower than a certain threshold are then filtered out. Results show that this method is reliable for terrain scenes where the object of interest covers between 5% and 50% of the scene [3]. In addition, for extremely cluttered scenes, the algorithm increases point-on-target selection by a factor of 2 to 7 relative to the remaining clutter points. The primary advantage of implementing the 3-D Cueing algorithm over alternative algorithms is that it is based on the same model representation scheme as Andrew Johnson's spin-image recognition method. The model representation overlap makes 3-D Cueing very attractive and economical to implement for our purposes.

1.2 Thesis Overview

From among the presented detection and recognition techniques, we believe that the spin-image based detection and recognition algorithms are the most promising for processing 3D Ladar terrain data. So far, the spin-image recognition method has been applied to a limited class of objects with simple geometries, such as rubber ducks, bunnies, toy robots and various pipe fixtures [25]. In addition, the processed scene data had high range resolution, and low sensor noise. The purpose of this thesis is to improve on the spin-image method for the detection and recognition of complex articulated objects in large terrain scenes, described by data with relatively low range resolution. The goal was to build an ATR system with a higher recognition rate (over 99%) along with a lower false alarm rate (less than 0.01%) as compared to current ATR systems [28] [29]. The rest of this paper is divided into three main areas: data preprocessing, target recognition and target detection.

Before proceeding into a description of the preprocessing, recognition and detection algorithms, an overview of spin-image surface matching is presented in Chapter 2. The overview is meant to provide a context for understanding the development of algorithms to follow.

1.2.1 Data Preprocessing

The raw data produced by the MIT Lincoln Laboratory Ladar sensor are several tri-dimensional angle-angle-range images. The target detection and recognition methods are not tuned to process this raw angle-angle-range data, but rather need a surface representation of the data. [24]. One convenient surface representation of the scene data is a 3D oriented data set, composed of 3D points along with their associated surface normal direction. The following processing steps were taken to obtain a 3D oriented data set from the raw angle-angle-range data:

1. Coordinate transformation from angle-angle-range to xyz
2. Data Filtering (Range Coincidence Processing)
3. Data Registration of the multiple-view xyz data frames
4. Surface Reconstruction of the registered xyz data

The scene pre-processing algorithms and results are discussed in Chapter 3 of this thesis.

In addition to converting the scene data to the spin-image representation, we also need a process to convert target CAD models into oriented 3D point data sets. We implemented a modeling procedure to generate high-resolution oriented 3D point-data sets from CAD models. To reduce online recognition time, spin-image libraries were generated offline from the 3D oriented data sets of the target models. Based on a visual analysis of clutter and occlusion in our particular measurement scenes, an optimal set of spin-image generation parameters is selected in order to maximize recognition performance while meeting our recognition run-time goals. Chapter 4 describes the implemented object modeling procedure, presents the resulting spin-image model libraries and discusses the selection of the optimal spin-image generation parameters for target detection and recognition.

1.2.2 Target Recognition

Chapter 5 describes the automatic target recognition algorithm that we implemented. The implementation relies on an augmented version of the spin-image surface matching process described in Chapter 2. Target recognition is demonstrated for twelve measured data scenes with targets both out in the open and under heavy canopy and camouflage cover. Correct target identification is demonstrated for targets with multiple movable parts that are in arbitrary orientations. Recognition quality and time performance results are presented and discussed.

1.2.3 Target Detection

Chapter 6 describes the implemented 3D Cueing target detection process combined with the target recognition process described in Chapter 5. The entire Automatic Target Detection and Recognition system is demonstrated on five measured scenes with targets both out in the open and under heavy canopy cover, where the target occupied between 1 to 10% of the scene by volume. Target detection and recognition results are presented and discussed.

A summary of the significance of the results obtained and directions for future work is presented in the Chapter 7 (Conclusion).

Chapter 2: Overview of Spin-Image Surface Matching

2.1 A Pose-Independent Representation of Surface Shape

In the spin-image based representation, surface shape is described by a collection of oriented points, 3-D points with associated surface normals. In addition, each 3D oriented point has an associated image that captures the global properties of the surface in an object-centered local coordinate system [24, pg.3]. By matching images, correspondences between surface points can be determined, which results in surface matching. Figure 2-1 depicts the surface-matching concept.

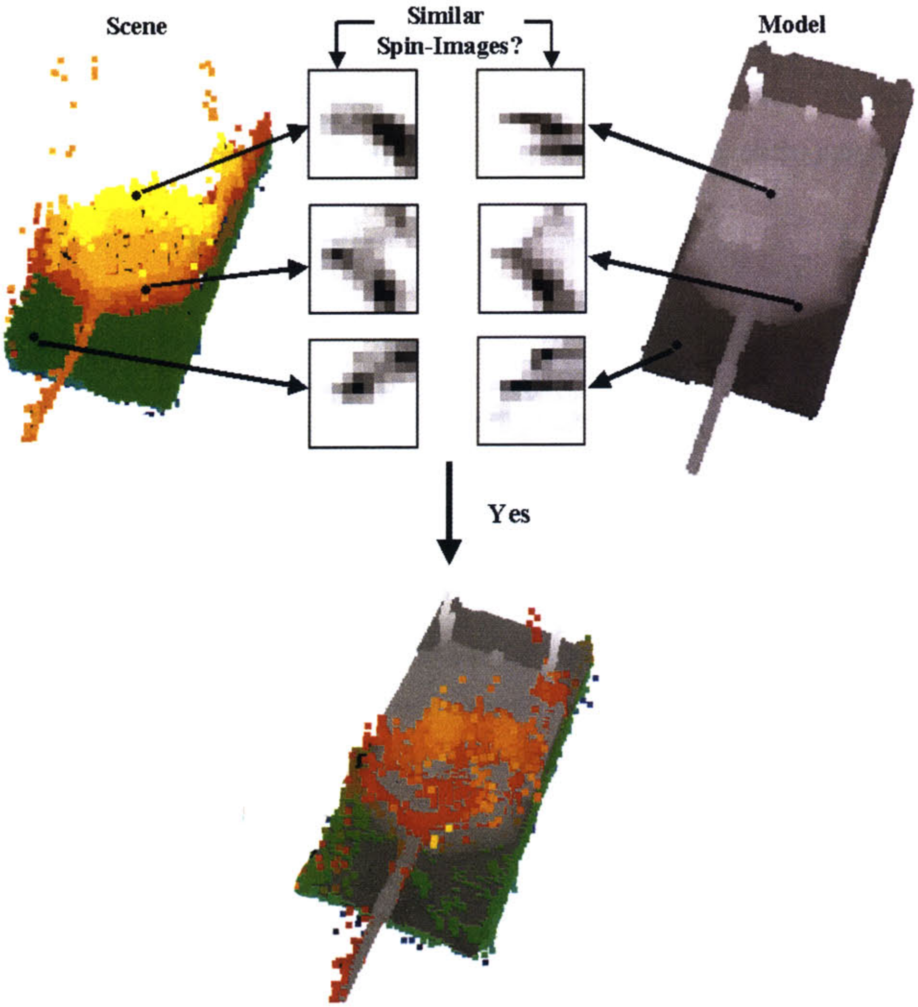


Figure 2-1: Spin-image Surface-Matching Concept

The image associated with each 3D oriented point is known as a spin-image. A spin-image is created by constructing a local basis at an oriented point. Using this local coordinate system, the position of all the other points on the surface can be encoded by two parameters. By mapping many of the surface points to this 2D parameter space, a spin-image is created at each oriented point. Since a spin-image encodes the coordinates of the surface points with respect to a local coordinate basis, it is invariant to rigid 3D transformations. Given that a 3D point can now be described by a corresponding image, we can apply robust 2D template matching and pattern classification to solve the problem of surface matching and 3D object recognition. [24]

2.2 Spin-Image Generation

The fundamental component for creating a spin-image is the associated 3D oriented point. As shown in Figure 2-1, an oriented point defines a 5-degree of freedom basis, using the tangent plane P though point p, oriented perpendicularly to normal N.

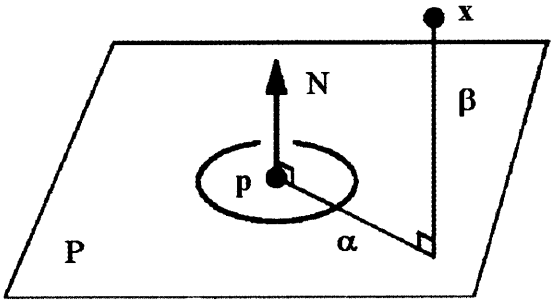


Figure 2-2: An oriented point basis created for a 3D point p.

Two coordinates can be calculated given an oriented point: α the perpendicular distance to the surface normal N and β, the signed perpendicular distance to the plane P. [24] Given an oriented point basis O, we can define a *Spin-Map* function that projects 3D points x to the 2D coordinates of a particular basis (p,n) as following:

$$So : R^3 \rightarrow R^2$$

$$So(x) \rightarrow (\alpha, \beta) = \sqrt{| x - p |^2 - (n \cdot (x - p))^2}, n \cdot (x - p)) \quad \textbf{(2.1)}$$

Applying the function So(x) to all the oriented points in the 3D point cloud will result in a set of 2D points in α–β space. To reduce the effect of local variations in 3D point positions, the set of 2D points is gridded to a 2D array representation. The procedure to create a 2D array representation of a spin-image is described visually in Figure 2-3. To account for noise in the data, the contribution of a point is bilinearly interpolated to the four surrounding bins in the 2D array. By spreading the contribution of a point in the 2D array, bilinear interpolation helps to further reduce the effect of variations in 3D point position on the 2D array. This 2D array is considered to be the fully processed spin-image.

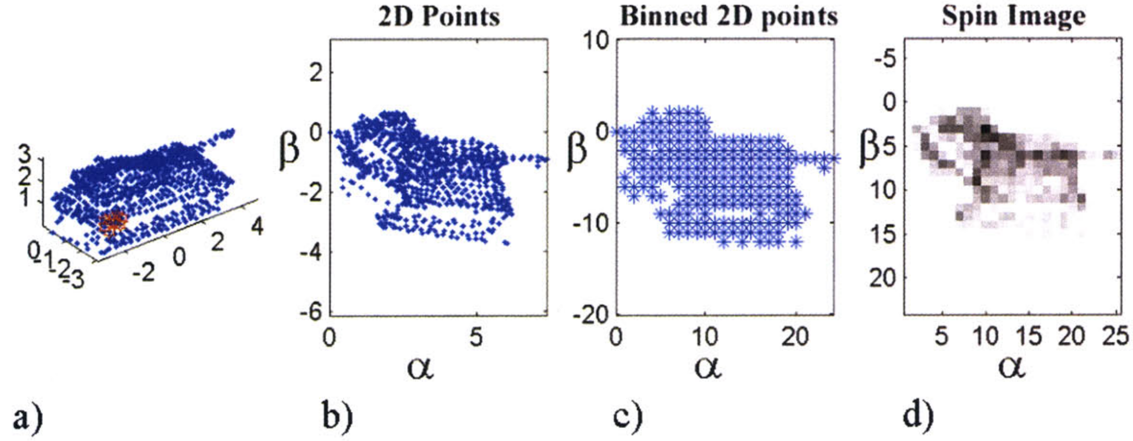


Figure 2-3: 2D Array representation of a spin-image using bilinear interpolation. a) Measurements of an M60 tank. Red dot indicates the location of the 3D point used to create the example spin-image. b) Resulting mapping of the scene points in the α–β Spin-Map of the chosen point, c) Spin-Image showing the non-zero bins after applying bilinear interpolation, d) Spin-Image showing the bin-values on a gray color scale. The darker bins indicate that a larger number of points were accumulated to those particular bins.

There are three parameters that control spin-image generation, namely bin size, image width and support angle. The bin size parameter determines the averaging in spin images that occurs during the process of binning the 2D spin-mapped points to the 2D array representation. According to A. Johnson, the bin size is set to a multiple of the data resolution; the acceptable range is between one to two times the data resolution. In this bin size range, the 2D bilinearly interpolated array adequately blurs the position of individual points while still maintaining a good description of the global surface shape. Figure 2-4 shows spin-images generated for a BMP Armored Personnel Carrier (APC) model using different bin sizes.

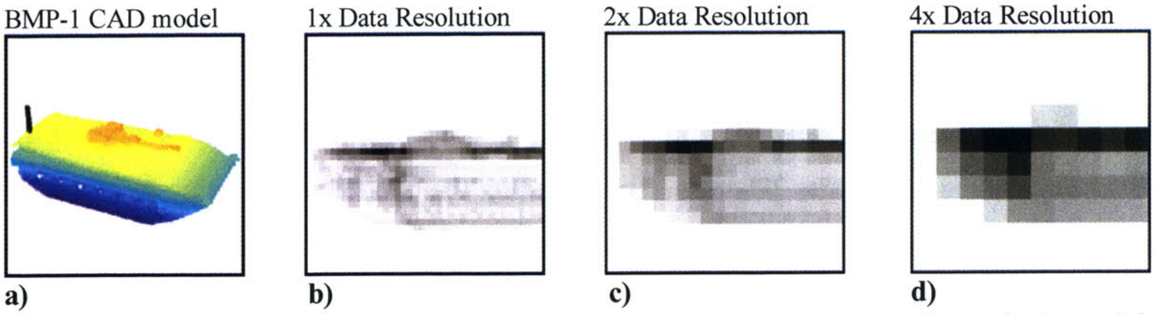


Figure 2-4: The effect of bin-size on spin-images. a) Height color-coded BMP CAD model. The black line corresponds to the normal of a 3D point. b) Three spin images created from that particular oriented point basis, with increasing bin-size.

The second parameter is the image width. When binning the 2D $\alpha-\beta$ points to the 2D array representation, the resulting spin-image can have any number of row or columns. For simplicity, a spin image is reduced to an equal number of rows and columns. This results in a square spin-image whose size is defined by one parameter, image width. Image width times the bin-size is defined as the spin-image support distance Ds. The support distance defines the dimensions of the space that can contribute points to a spin-image. By controlling the support distance, the amount of global surface information can be controlled [24, pg 25]. For the results presented in this thesis, the image width is set between 5 and 10, resulting in spin-images with 25 to 100 bins.

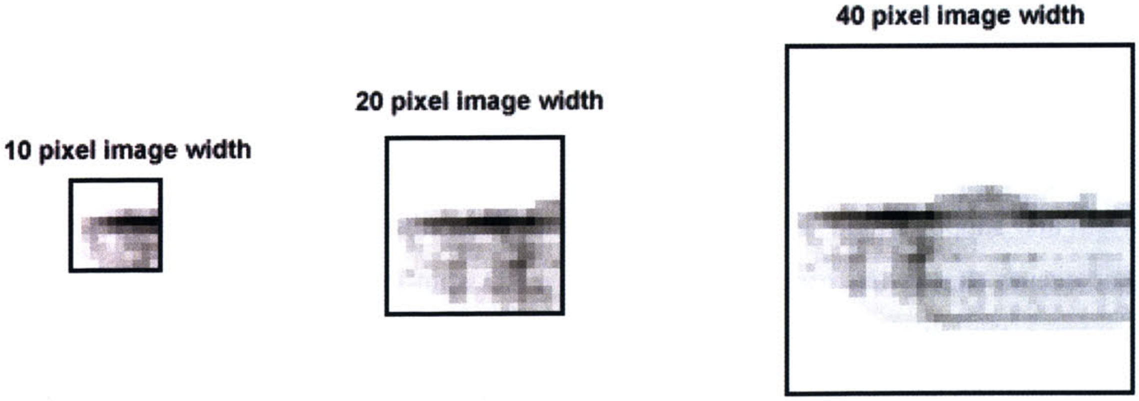
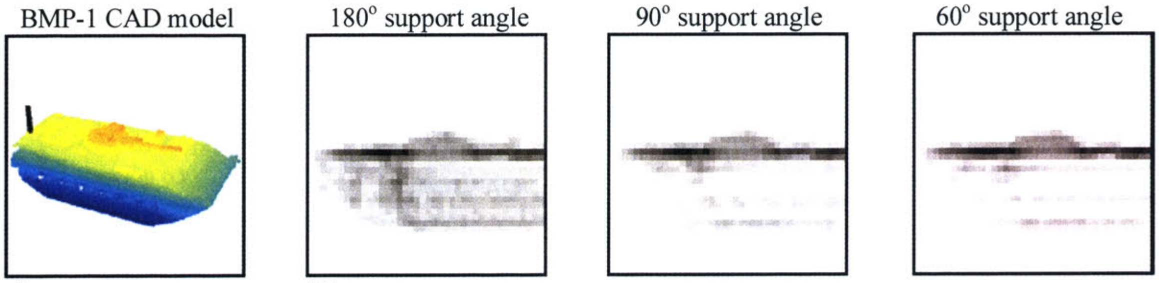


Figure 2-5: The effect of image width on spin-images. An increase in image width results in a proportional increase in support distance. Varying the image width allows spin-images to vary smoothly from local to global representations.

The third spin-image generation parameter is the support angle (As). The support angle is defined by A. Johnson as "the maximum angle between the direction of the oriented point basis of a spin image and the surface normal of points that are allowed to contribute to the spin-image." [24, pg 27] Let's say we have an oriented point A with position and normal (p_A, n_A) and an second oriented point B with position and normal (p_B, p_B). Then, B will be included in the spin-image of A if

$$a\cos(n_a \cdot n_b) < As \tag{2.2}$$

This parameter is useful in limiting the effect of object self-occlusion and nearby clutter during spin-image matching. Figure 2-6 shows the spin-images generated for three support angles for an oriented-point on a BMP-1 model.



a) b)

Figure 2-6: The effect of support angle on spin-images. a) Height color-coded BMP CAD model. The black line corresponds to the normal of a 3D point. b) Three spin images created from that particular oriented point basis, with decreasing angle of support.

2.3 3D Surface Matching

The implemented surface-matching algorithm closely follows the procedure described in Chapter 3 of A. Johnson PhD thesis. Given a scene and model data set, the sampling density of both data sets is first reduced to the same resolution by 3D voxel sub-sampling. The 3D voxel sub-sampling process has a similar effect as A Johnson's procedure for adjusting mesh-resolution. Normals are calculated based on the sub-sampled data set along with the rough 3D Ladar sensor position. (See Section 3.2.3.)

For each sub-sampled data set, a spin image stack is created. Each spin image in the scene data set is correlated to all the model spin-images, resulting in a distribution of similarity measure values. The correspondences obtained for each scene spin-image to model spin-image stack comparison are filtered using a statistical data based similarity measure threshold. The above process is repeated for the rest of the scene spin-images, resulting in a wide distribution of similarity measures.

Given the new distribution of similarity measures, a second similarity threshold is applied to remove unlikely correspondences. The remaining correspondences are further filtered and then grouped by geometric consistency in order to compute plausible transformations that align the scene to the model data set. The initial scene to model alignment is refined using a modified version of the Iterative Closest Point algorithm (ICP) in order to obtain a more definite match. Figure 2-7 below shows a detailed block diagram of the surface matching process.

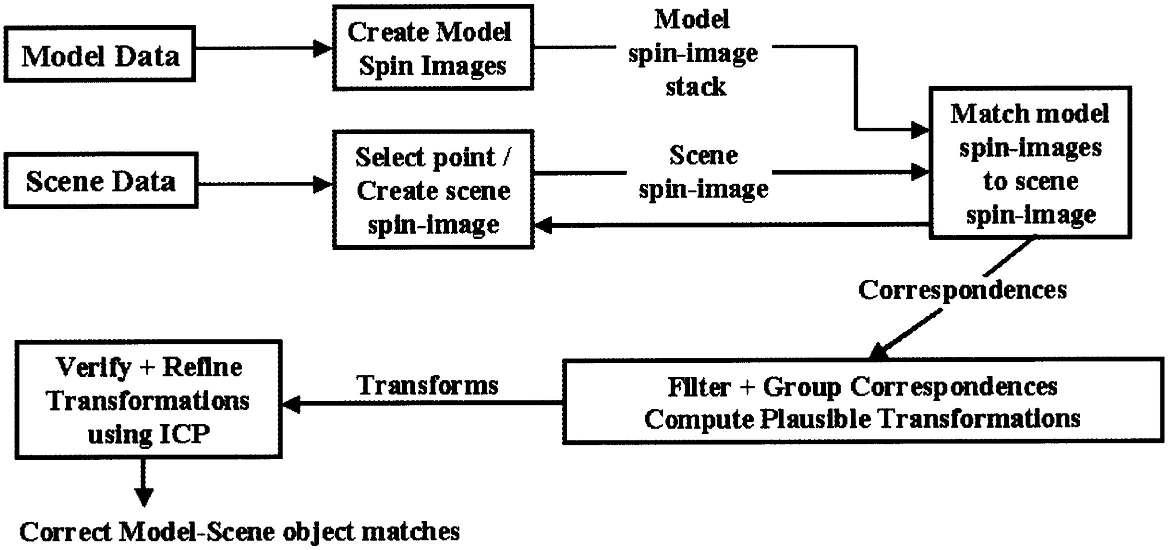

Model Data
Create Model Spin Images
Model spin-image stack
Scene Data
Select point / Create scene spin-image
Scene spin-image
Match model spin-images to scene spin-image
Correspondences
Verify + Refine Transformations using ICP
Transforms
Filter + Group Correspondences Compute Plausible Transformations
Correct Model-Scene object matches

Figure 2-7: Surface Matching Block Diagram

This particular surface-matching process is versatile since no assumptions are made about the shape of the objects represented. Thus, arbitrarily shaped surfaces can be matched, *without the need for initial transformations*. This is particularly critical for our target recognition problem where the target's position and pose is unknown. Furthermore, by matching multiple points between scene and model surfaces, the algorithm can eliminate incorrect matches due to clutter and occlusion. In the following sub-sections, we will explain in more detail the surface-matching process.

2.3.1 Spin-Image Matching

Suppose that we have a scene containing an instance of an object for which we have a complete object model. Furthermore, assume that the object in the scene has an unknown pose orientation. Since spin-images are pose-independent, given spin-images from surface of measured object in the scene, we expect to have model spin-images that are similar. By directly comparing scene to model spin-images, we should be able to find point correspondences between points on the surface of the scene object and points on surface of the complete object model.

Since spin-images from an instance of a model object will not be the same as the spin-images obtained from the measured object in the scene, we need to have a method to compare two spin-images. We expect that two spin-images created from proximal points on the surface of the model and scene instance of the object to be linearly related because the distribution of points falling in a corresponding bins will be similar (given that the model and scene data sets were sampled at the same resolution). A standard technique for comparing linearly related images is normalized correlation. Given two spin images P and Q with N bins each, the normalized correlation value R(P,Q) is

$$R(P,Q) = \frac{N\sum p_i qi - \sum p_i \sum q_i}{\sqrt{(N\sum p_i^2 - (\sum p_i)^2)(N\sum q_i^2 - (\sum q_i)^2)}} \qquad \textbf{(2.3)}$$

R ranges between –1(anti-correlated) and +1 (completely correlated). The function R provides a method to compare two spin-images: if R is high, then images are similar, while when R is low, the images are not similar.

An optimization is added to the computation of the normalized correlation that attempts to mitigate the effects of clutter and occlusion on a particular scene spin-image. The optimization is that N is defined as the number of bins for which both spin-images have data. Thus, only bins with data in both spin-images are considered in calculating the correlation coefficient. Preventing non-zero spin-image bins in the scene spin-image from matching zero-valued model spin-image bins mitigates the effects of clutter. That is, considering the generation of a spin-image, scene bins where the model indicates no data should exist are likely to be due to clutter and should not be considered. Conversely, preventing non-zero bins in the model spin-image from matching zero-valued scene spin-image bins mitigates the effects of occlusion. Again, based on spin-image generation, we can deduce that non-zero model bins that have corresponding scene bins with no data are likely due to an occlusion of the object surface in the scene. A. Johnson et al. clutter analysis results confirm the validity of this optimization.

As defined, the normalized correlation value does not take into account that a spin-image comparison with high bin overlap should have a higher confidence in the correlation result than a spin-image comparison with a low bin overlap. Since two spin-images with more overlap should be given a higher correlation value, a confidence measure is incorporated into the final spin-image similarity measure. One way to measure confidence in the correlation coefficient is to determine its variance. Thus the similarity measure incorporates the normalized correlation value along with its variance, and is implemented as follows:

$$C(P,Q) = (a\tan(R(P,Q)))^2 - \lambda(\frac{1}{\mathrm{N}-3}) \tag{2.4}$$

The similarity measure will return a high value for two images that are highly correlated and also have a large number of overlapping bins. The λ is a constant that is derived from the model spin-image stack bin occupancy. λ represents the expected overlap between spin-images. To determine λ, the bin-occupancy for each model spin-image is first computed. The bin-occupancy is the number of bins in a spin-image that have non-zero weight values. The median bin-occupancy value is found and λ is set to ½ of that particular value.

2.3.2 Correspondence Filtering

For each selected scene point, a spin-image is created. The scene spin-image is then correlated to all the model spin-images, resulting in a distribution of correlation values that has a single mode corresponding to incorrect spin-image matches [24]. The highest outliers in the distribution represent correct spin-image matches. For single-mode distributions, a standard way to determine statistical outliers is to compute the fourth spread of the histogram (fs = upper fourth – lower fourth = median of largest N/2 measurements - median of smallest N/2 measurements). Statistical extreme outliers are defined as 3 fs units above the upper fourth. These statically extreme outliers are picked out of the correlation-value distribution and marked as valid correspondences.

The above correlation process is repeated for a sampling of all scene data points. The sampling ranges from 20% to 50% of all scene data points. Scene data points are not judiciously selected: the sample points are uniformly distributed across the given scene. Therefore, no feature extraction was performed to pick certain points. Given all the found correspondences, several filtering methods were performed to remove unlikely correspondences.

The first filtering step uses similarity measure to remove unlikely correspondences. A similarity measure threshold is applied to the new distribution of correspondence similarity measures. The similarity threshold is defined as a fraction from the maximum similarity measure value. The fraction was set to 0.4. In addition, in order to prevent combinatorial explosion for large scene data sets, a maximum of 4000 correspondences were kept after the fraction-of-max threshold was applied. Thus, if more than 4000 correspondences are above the fraction-of-max threshold, the similarity measure values would be used to select the highest 4000 correspondences. By setting the threshold to a fraction-of-max, we are certain to threshold out any similarity measures that are smaller than 0, which represents an uncorrelated result. This threshold is reasonable since we are not interested in correspondences that range from anti-correlated to uncorrelated. Thus, this filtering step has the potential to filter out all the correspondences. Given such an event, the surface matching process is terminated with an indication that no match was found.

The second filtering step uses geometric consistency to remove unlikely correspondences. Geometric consistency estimates the probability that a set of two correspondences can used to calculate a good transformation in order to align the model to the scene [24]. The geometric consistency of a set of two correspondences is measured in the spin-image space of the corresponding points. The geometric consistency for correspondence $C_1=[s_1,m_1]$ to $C2=[s_2,m_2]$ is

$$d_{gc}(C_1,C_2)=\frac{|S_{m2}(m_1)-(S_{s2}(s_1)|}{(|S_{m2}(m_1)|+|(S_{s2}(s_1)|)/2}$$
$$D_{gc}=\max(d_{gc}(C_1,C_2),d_{gc}(C_2,C_1)) \qquad \textbf{(2.5)}$$

D_{gc} measures the distance between correspondences in spin-image coordinates, normalized by the average vector length of the spin-image coordinates, in order to prevent a preference towards correspondences that are near each other. Because d_{gc} is not symmetric, the maximum of $d_{gc}(C_1,C_2)$ and $d_{gc}(C_2,C_1)$ is used to define D_{gc}, the geometric consistency distance: as a result, if points are geometrically consistent, D_{gc} will be small.

Given as set of correspondences that has been filtered by similarity measure to a list L of correspondences, we follow A. Johnson's procedure to determine if a correspondence C_1 is geometrically consistent. First, a threshold T_{gc} is set such that if $D_{gc}(C_1, C_2) < T_{gc}$, then C_1 is geometrically consistent to C_2. According to A. Johnson, for high geometric consistency, the optimal T_{gc} is set to 0.25. For each C_1 in L, $D_{gc}(C_1, C_2)$ is computed for all the other C_2 in L. If more than a quarter of the number of correspondences in L are geometrically consistent to C_1, then C_1 is considered geometrically consistent. The above steps are applied to the rest of the correspondences in L. Only geometrically consistent points are kept. In our tests, the number of correspondences left is between 20 to 800 correspondences. Similar to the similarity measure threshold, the geometric consistency threshold has the ability to filter out all the correspondences, resulting in the return of a no-match. Next, the filtered correspondences are grouped into sets that can be used to compute model to scene transformations.

2.3.3 Pose Estimation and Verification

The remaining correspondences are grouped based on the criterion W_{gc}, which is similar to D_{gc}, with an added weight that encourages grouping of correspondences that are far apart.

$$
\begin{aligned}
w_{gc}(C_1,C_2) &= \frac{d_{gc}(C_1,C_2)}{1-e^{-[(|S_{m2}(m_1)|+|(S_{s2}(s_1)|)/(2\gamma)]}} \\
W_{gc} &= \max(w_{gc}(C_1,C_2),w_{gc}(C_2,C_1))
\end{aligned}
\qquad \textbf{(2.6)}
$$

The value of W_{gc} will be small when two correspondences are geometrically consistent and also far apart. The grouping equation is based on geometrical consistency because geometrically inconsistent points will produce highly erroneous transformations. In addition, the grouping equation picks points that are spread further apart from each other since points that are close together generate transformations that are vulnerable to noise in point position [30]. The constant γ is a scale-independent normalization weight that promotes the grouping of points that are far apart. The value of γ is set to four times the data resolution in order to prevent correspondences that are closer than 4 times the data resolution from being grouped together.

Given the previous list L, grouping for each correspondence in the list is performed as follows:

1. Select a seed correspondence Ci in L and create a group Gi={Ci}
2. Find a correspondence Cj that has the minimum W_{gc}(Cj,Gi) value.
3. If the W_{gc}(Cj, Gi) < T_{gc}, then add Cj to the the group. T_{gc} is set to 0.25.
4. Repeat steps 2 and 3 for all the remaining correspondences.

After applying the grouping algorithm, we will have groups that have anywhere from one to about twenty correspondences. Groups with less than three correspondences are considered weak groups and are thrown out. From each remaining group, a rigid transformation T from model to scene is calculated by minimizing the least-squares error [31]:

$$E_T = \sum | s_i - T(m_1) |^2 \tag{2.7}$$

Given these groups and associated pose transformations, we can now verify the pose transformations in order to eliminate any inconsistent matches between the scene and the model data. In practice, the number of groups ranges from 50 to 250. Since it would take a long time to fully verify each correspondence group, we do an initial verification for each group in order to find the most plausible pose transformations. If any plausible pose verifications are found, a certain number of transforms that are the most promising are then fully verified.

The verification algorithm is based on a modified version of the Iterative Closest Point (ICP) algorithm, which can handle partially overlapping surfaces. [32, 33, 34]. The ICP algorithm iteratively determines a transformation between a scene and a model by creating point-pair correspondences, applying the newly found transform to all the model points and then repeating the process. Figure 2-8 shows a block diagram of a generic version of ICP.

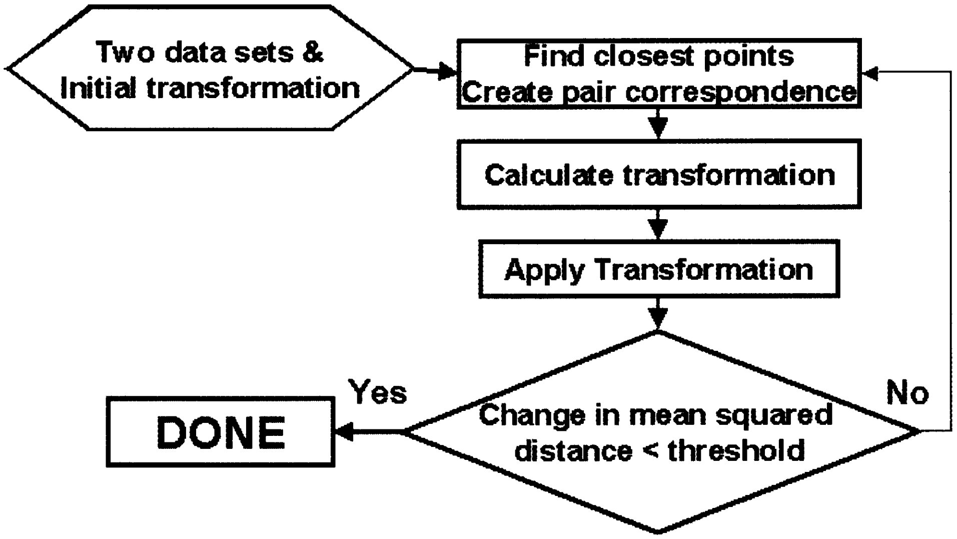


Figure 2-8: Block Diagram of Generic ICP

One of the drawbacks of ICP is that it converges to a local minimum in pose-distance space. Therefore, the initial position of the two data sets is crucial: while the algorithm works well for small transformations, the performance can degrade for arbitrarily large transformations. Another problem with the generic form of ICP is that it assumes that one data set is a subset of the other, with complete overlap to the larger data set. In practice, our scene is not a subset of the model, since it might include clutter and noise that is not present in the model. Our verification algorithm is based on a modified ICP that can handle partially overlapping surfaces. In addition, the modified ICP algorithm also takes as input an arbitrary transformation based on the initial spin-generated pose transformation that roughly aligns the scene and model. By applying the spin-image generated transformation, only a small transformation error remains that needs to be corrected by ICP.

Since the two data sets partially overlap, a method is needed to limit the creation of point correspondences only to those areas in the two data sets that overlap. Our verification algorithm does this by creating 3D search voxels centered at each model point. The Voxel size is set to 2 times the data resolution. The resulting voxel is searched for scene points. If no scene points are found, then that particular model point is considered not to overlap with the scene. If one or more scene points are found, the closest scene point to the respective model point is picked to create a point correspondence.

The closest point distance between a scene and a model point is defined by the 3D position and surface normal. Given two oriented points (s_p, s_n) and (m_p, m_n), the closest point distance is defined as

$$D=\sqrt{| s_p - m_p | + w_n | s_n - m_p |} \qquad \textbf{(2.8)}$$

,where w_n is the weight of normals as compared to surface position. W_n is set to one times the scene resolution. To improve the speed for the closest point search, a six-dimensional k-D tree is created for the model data set. Using the model data set k-D tree, scene points from overlapping surfaces can be found efficiently. [34]

Each spin-image generated pose transformation is initially verified by running one iteration of ICP. If the spin-image pose transformation correctly aligns the model to the scene, then ICP should report a high overlap and low mean-squared error (MSE) after one iteration. A pose transformation is considered plausible if the transformed model has more than 10% overlap with the given scene. Given a plausible pose transformation, the quality of the pose transformation is determined based on a goodness of fit value defined as follows:

$$GOF = \frac{\theta^2}{MSE} \tag{2.9}$$

θ is the fraction of overlap between the scene and the model. θ is defined as:

$$\theta = \frac{S}{M} \tag{2.10}$$

, where S is the number of scene to model correspondences found and M is the number of model points. A higher GOF indicates a higher fraction of overlap and/or smaller MSE, thus an increased likelihood that the pose transformation correctly aligns the model to the scene. The above verification process is applied to all the spin-image pose transformations. Based on the GOF value, the best 25 pose transformations are picked for full verification.

Similar to the initial verification, the full verification process takes a plausible pose transformation and runs the ICP algorithm in order to refine the pose transformation. For full verification, ICP is run for a maximum of 50 iterations. Based on observations, this number of iterations should be sufficient to correctly align plausible pose transformations to within less than 1^0, given a pose error of at most $30\text{-}45^0$ in roll/yaw/pitch. A GOF verification value is computed for each pose transformation. The Verification GOF (V_{GOF}) is defined as:

$$V_{GOF} = \frac{(\theta^2 \cdot N_{pt})}{MSE} \tag{2.11}$$

, where N_{pt} is the number of plausible pose transformations found. The number of pose transformations is included in the V_{GOF} value because a higher number of pose transformations indicates a higher level of confidence that the model matches the scene.

2.4 Discussion and Summary

The spin-image representation is useful for several reasons:

1. Pose Invariance allows one to uniquely describe an instance of that object regardless of its pose. This feature of spin-images allows the creation of a compact representation that breaks the problem of 3D recognition into a set of 2D recognition problems.
2. Has minimal requirements on surface shape.
3. Can smoothly scale from a local to a global representation through the adjustment of image width.
4. Is robust to clutter and occlusion through the use of support angle and the 2D normalized correlation on only non-zero spin-image bins.

The spin-image representation is well tailored to our particular target recognition problem. The task of our particular sensor is to identify targets under heavy camouflage and canopy cover. Therefore, the resulting data sets will have a lot of scene clutter around the target and relatively large levels of target occlusion. The spin-image representation can readily handle both target occlusion and clutter by controlling the respective spin-image parameters, namely support angle and image width.

Aside from the benefits of the spin-image representation, the spin-image matching process attempts to discriminate and correctly address most, if not all of the typical matching scenarios, namely:

1. The scene does not contain a target object
2. The scene contains an unknown target object
3. The scene contains a known target object
4. The scene contains multiple known/unknown target objects

One important scenario is Case 1, where the correct answer is “none of the above.” Case 1 from above is addressed in several ways by the surface-matching process (see Figure 2-7 for an overview of the process of surface matching):

1. One possibility for correctly returning a "no-match" result is that the similarity measure filter will find completely uncorrelated results. As a result, no correspondences will be left and a "no match" will declared.
2. If some correspondences do pass the similarity measure filter, the geometric consistency filter is very likely to return a "no-match" result because, as a group, the scene clutter correspondences are extremely unlikely to have the same relative geometric positions as the points on the currently-tested model object.
3. If by chance some of correspondences pass the geometric consistency filter, the need to create a self-consistent correspondence group will place further constraint on the remaining correspondences. If groups of less than 3 correspondences are created, then a "no-match" is again returned.
4. A further constraint is placed by the selection of plausible transformations. If no plausible transformations are found, a ”no match" is again returned.
5. If the above constraints are passed, we will get a match. However, the match is bound to have a low GOF value, indicating a low likelihood of a correct match. At this stage, we might still be able to rectify an otherwise incorrect result by setting an arbitrary threshold on the GOF value to prevent false matches.

A Case 2 scenario, where the scene contains an unknown target object, is handled similarly to Case 1. However, depending on the similarity of the object to a known target object, it is possible that the unknown target might be considered a match based on its higher-level features; thus, if the unknown object is the BMP-1 Armored Personnel Carrier (APC), it might match the similarly shaped APC, the BMP-2 (see Figure 4-1 for images of the two targets). While some purists might argue that a BMP-2 model matching a BMP-1 scene target should be considered a incorrect match, it is actually very useful to classify objects based on their general features because we can greatly reduce both the recognition time and the size the object model library. Furthermore, if the surface matching process relies on the general surface features of a target, then the matching algorithm will probably be more resistant to measurement noise and any relatively small changes in the configuration of the target. (i.e. the mounting of a machine gun on top of the target's roof, the placement of extra fuel barrels on a tank). Being able to correctly classify targets based on the general target shape is especially important for the recognition of military ground vehicles, where the vehicles of interest come in many possible configurations.

For Case 3, where a known target object is in the scene, we expect to obtain a match that correctly brings into alignment the scene target with the model object. Given ideal targets with no noise, clutter or occlusion, the surface-matching algorithm is almost certain to recognize the target and find the correct pose. However, correct recognition gets progressively harder with an increase in the amount of scene noise, scene clutter and target occlusion. Since we expect that most of our recognition scenes will have some measurement noise along with relatively high levels of nearby clutter (i.e. clutter within 1 meter proximity to the target representing up to 50% of the data set) and occlusion of up to 70%, we need a recognition algorithm that can handle these particular issues. The effects of scene noise, clutter and occlusion are addressed in several ways at each step of the spin-image surface matching process:

1. In its essence, scene noise affects the relative 3D position of a point relative to the position of the rest of the scene points. In our spin-image representation, the uncertainty in the 3D position of a point translates to uncertainty in the $\alpha-\beta$ spin-map projection. The uncertainty in $\alpha-\beta$ position is addressed by bilinearly interpolating the $\alpha-\beta$ 2D points to a 2D array. The 2D bilinearly interpolated array adequately blurs the position of individual points while still maintaining a good description of the global surface shape. Thus, the resulting spin-image will not be sensitive to small changes in the 3D position of a point.
2. The problems of clutter and occlusion are mitigated when computing the similarity measure by allowing only bins with data in both the model spin image and the scene spin-image to be considered in calculating the correlation coefficient. Furthermore, the similarity measure threshold uses a data-based threshold, which will allow correct correspondences to pass regardless of the number of correspondence found, a figure that might depend on the amount of

occlusion or lack of it in the scene. Thus the similarity measure filter should not be greatly affected by the presence of clutter or occlusion.

3. The geometric consistency filter again is based on a threshold that is relative to the current data set (i.e. 25% of found correspondences must be geometrically consistent with the current correspondence in order for the particular correspondence to pass the filter). By using a relative threshold, geometric consistency will not be affected by occlusion, which directly affects the number of points on the target surface, and therefore the number of possible correspondences to be found on the target surface.
4. The computation of plausible transformations places constraints that are minimally affected by occlusion and clutter. The constraints imposed to determine a plausible transform are that we have at least three correspondences and at least 10% coverage of the target. In order to solve a 3D absolute orientation problem, we need at least three correspondences, hence the first requirement on the number of correspondences. The second requirement on target coverage is a relatively weak constraint needed to prevent false alarms. Given that the scene object is no more than 90% occluded, it is theoretically possible to find a plausible transformation that correctly aligns the surface of the scene target measurement to the surface of the model object.

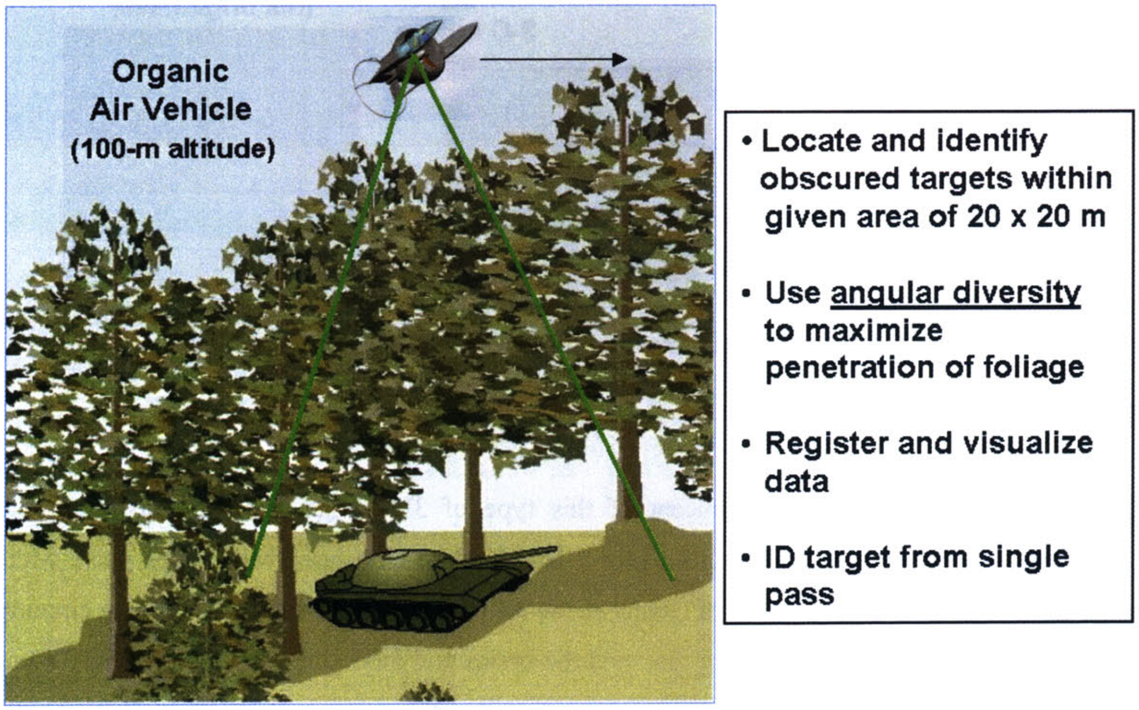
The last case is a scene that contains multiple known/unknown target objects. This case can be trivially handled by the spin-image representation, which can smoothly scale from a local to a global representation through the adjustment of image width. By adjusting the image width parameter, the spin-images from one particular target will be unaffected by the presence of nearby targets or clutter. The end-result is that each target in the scene will be just as likely to match a known object, regardless of presence of nearby targets or clutter. Thus, the problem can be broken down into several independent "Case 2" and "Case 3" scenarios and handled appropriately as described above.

In summary, the spin-image representation and spin-image matching algorithm has the potential to be widely applicable to problems in 3D computer vision. In the next chapter, we discuss data acquisition and pre-processing methods needed to reconstruct a scene given measurements from multiple viewing directions. The reconstructed scene can then be converted into the spin-image representation in order to perform surface matching. Further algorithmic augmentations to the current spin-image matching process lead to the development of an Automatic Detection and Recognition system described in Chapter 5 and Chapter 6.

Chapter 3: Acquisition and Preprocessing of Pulsed 3D Ladar Imagery

3.1 Data Acquisition

The data was acquired with the JIGSAW sensor, which is an airborne platform designed to augment an UAV with 3D Ladar capabilities. Given a target cue from a large area ground search, the JIGSAW sensor flies over the designated location while constantly adjusting its pointing optics to track the respective ground region. Multiple perspectives of the region are taken in order to better reconstruct the scene and alleviate obscuration due to canopy and camouflage cover. Figure 3-1 graphically shows the JIGSAW concept along with the JIGSAW system goals.



- Locate and identify obscured targets within given area of 20 x 20 m
- Use angular diversity to maximize penetration of foliage
- Register and visualize data
- ID target from single pass

Figure 3-1: JIGSAW data acquisition concept and system goals.

3.1.1 JIGSAW 3D-Lidar Sensor

Figure 3-2 shows the ladar sensor concept. Light from a pulse laser is diverged to illuminate the scene of interest. The light reflected from the scene is detected onto a two-dimensional array of detectors. The detectors measure the relative time of arrival of the reflected light, rather then measuring intensity. The time of arrival is linearly dependent on the range from the detector array to the measured scene. Thus, the time of arrival data from each detector pixel can be used to produce an angle-angle-range, or tri-dimensional image from each laser pulse. This kind of ladar sensor can be used to penetrate foliage and identify obscured targets [35].

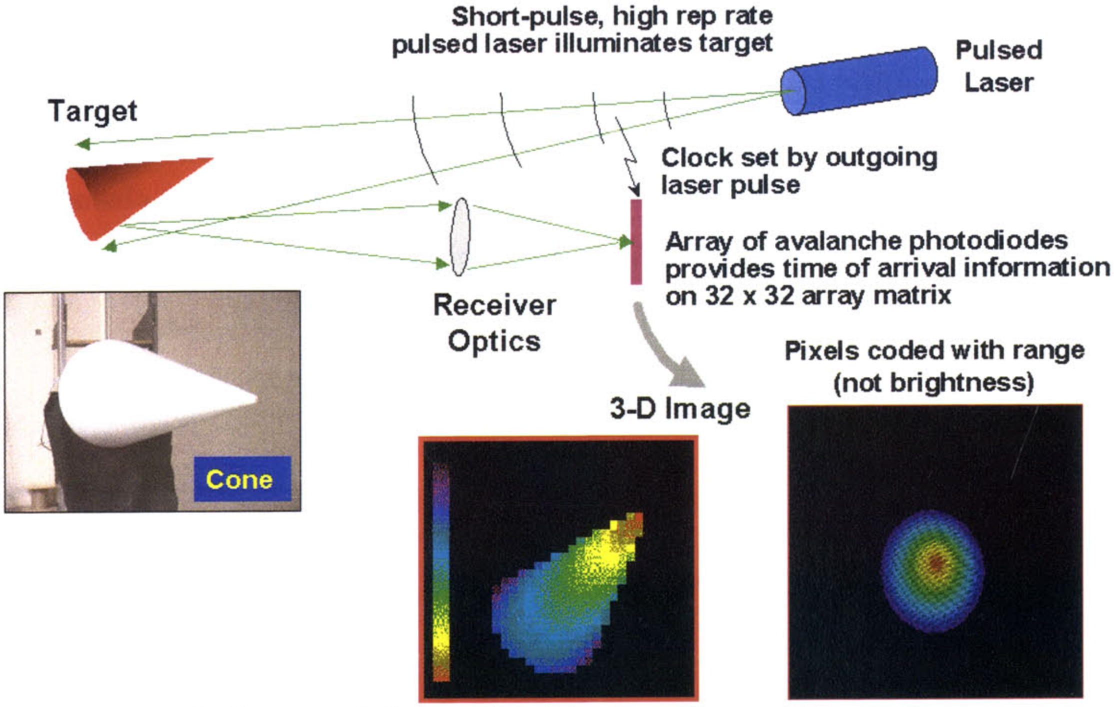


Figure 3-2: Basic Concept of tri-dimensional (angle-angle-range) laser radar.

3.1.2 Data Collection

To demonstrate the utility and concept of this type of 3D Ladar, Lincoln Laboratory has constructed the JIGSAW ladar system using a 16khz micro-chip laser with 300ps pulse width operating at 532 nm, a 32x32 array of Geiger-mode APDs integrated with CMOS timing circuitry providing a 2GHz effective sampling rate, and the optics and mounting to allow the entire scene to be scanned, building up high-resolution images with several hundred pixels in each angular direction. The current JIGSAW system can be mounted in a helicopter to collect data from a 150-450 meter altitude above a target of interest.

3.2 Data Processing

The raw angle-angle-range data obtained from our ladar sensor is subjected to several processing steps. The following processing steps were implemented to obtain a measurement of the target that can be visualized in the spin-image representation and identified using spin-image matching:

- Coordinate transformation from angle-angle-range to xyz
- Data Filtering (Range Coincidence Processing)
- Data Registration
- Surface Reconstruction

Coordinate transformation from sensor angle-angle-range space to world xyz space is performed using an APPLANIX 3.0 GPS/INS system. The INS sub-system records yaw, roll and pitch in order to determine the platform orientation and orientation rate of change. The GPS sub-system records the current platform position, in order to account for the translation of the sensor.

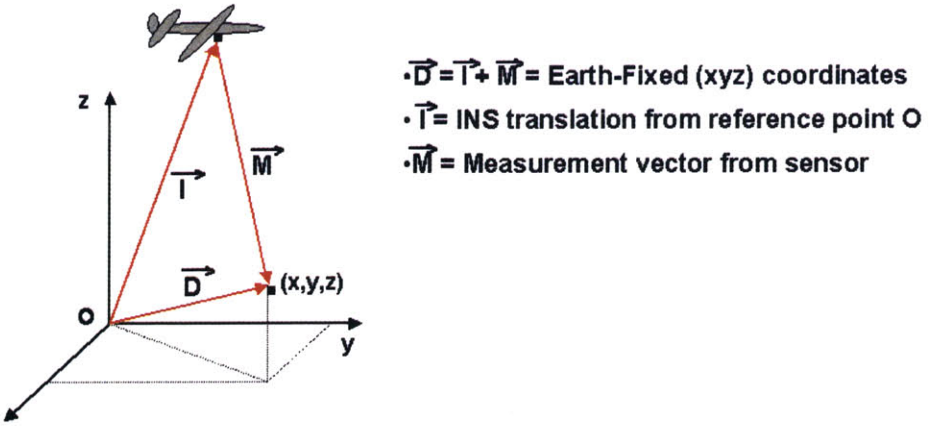
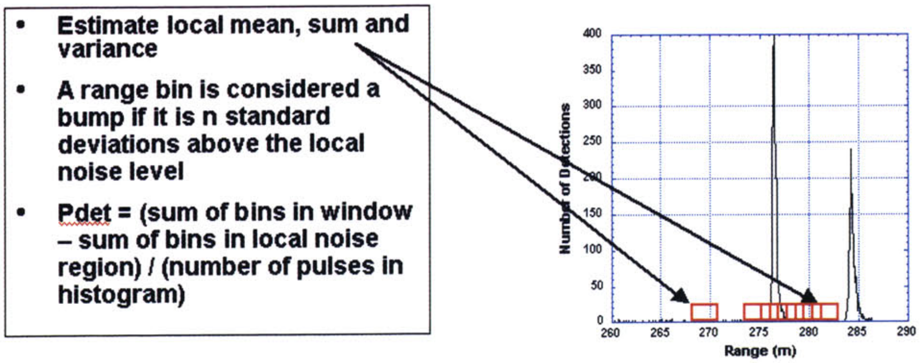

z
$\vec{M}$
$\vec{I}$
$\vec{D}$
(x,y,z)
O
y
x
•$\vec{D}=\vec{I}+\vec{M}$ = Earth-Fixed (xyz) coordinates
• $\vec{I}$= INS translation from reference point O
•$\vec{M}$ = Measurement vector from sensor

Figure 3-3: Coordinate transformation from sensor-angle-angle coordinates to Earth-Fixed coordinates. $\underline{M}$ represents the raw range coordinates, $\underline{I}$ represents the INS translation vector and $\underline{D}$ represents the resulting Earth-Fixed xyz coordinates.

3.2.1 MPRCP Algorithm

The raw sensor data are now in the Earth-Fixed coordinates. Typically, these raw data have a large amount of noise, most of which is present in the original range direction.

This range noise is more commonly known as a range tail. In order to filter out the range tail, we need to apply a filter in the range direction. The particular filter should remove all the data except the most significant ranges. The MPRCP (Multi-Peak Range Coincidence Processing) algorithm was implemented to filter out range tails. The MPRCP concept is pictorially shown below:



- **Estimate local mean, sum and variance**
- **A range bin is considered a bump if it is n standard deviations above the local noise level**
- **Pdet = (sum of bins in window – sum of bins in local noise region) / (number of pulses in histogram)**

Figure 3-4: A typical range histogram showing multiple range peaks corresponding to hard target surfaces, such as trees and the ground.

The MPRCP algorithm takes as input a frame of data in xyz space and temporarily transforms it to sensor-view angle-angle-range space. Based on the sensor's particular angle-angle resolution setting, the data set is binned into a 2D array of range histograms. Each range histogram contains all the range hits in that particular range angle-angle bin. The range hits are binned using a range resolution of 7.5 cm, resulting in a histogram as shown in Figure 3-4. The range histogram is analyzed using a moving window, depicted in Figure 3-4 as a red box. For each window position, the local mean, sum and variance are calculated. A range window is considered to have significant range values if it is *n* standard deviations above the local noise level. The moving window is applied across the entire histogram, and only the significant range peak locations are kept. In addition, a probability of detection Pdet is calculated for each data point as follows:

$$P\det = \frac{\sum Range\,Hits\,in\,window - \sum Range\,Hits\,of\,local\,noise}{\sum Range\,Hits\,in\,Histogram} \tag{3.1}$$

3.2.2 Registration Algorithm

Based on the performance of our particular GPS sensor and platform air speed, we were able to accumulate data in ¼-second acquisition intervals without any significant intra-alignment errors. Over a sequence of ¼-second MPRCP processed frames, the coordinate system of the processed data slowly drifts due to errors in GPS information. For targets measured from multiple views, the GPS translation errors result in a large misalignment of the measured target surface. For our particular GPS sensor, the largest drift error occurs in the height(z) direction. For a typical data collection of 50-100 ¼-second frames, the observed drift in the z-direction can be as much as 10 meters. The GPS error in the xy direction is smaller, ranging from 2 to 5 meters. Registration is used to align the measurements of the target surface in order to reconstruct a more complete 3D signature of the target of interest.

We developed a registration algorithm that corrects for this drift in a two-step process: First, the error in the z direction is corrected by detecting the location of the measured ground. The procedure for ground detection and registration is based on the spin-image matching technique discussed in Chapter 2. After the drift in the z direction is removed, the xy drift is corrected using the ICP algorithm previously described in Section 2.3.3.

3.2.2.1 Ground Detection and Registration

For a typical ¼-second measurement of a target under heavy tree cover, measured ground values account for only 5-10% of the total data set. We decided to use the spin-image matching algorithm because it is robust to such high levels of clutter within a scene.

Typically, the measured ground in our data collections is mostly flat and can be accurately modeled as 10x10 meter flat ground patches. Thus, a 10x10m flat plane defines the model that we are searching for in each processed frame. Using the spin-image matching procedure described in 2.3.1, correspondences are found between the model plane and the measured scene. In order to remove unlikely correspondences, the similarity measure filter is applied (See 2.3.2). For robust ground detection, the similarity measure threshold is set to ¼ of the maximum similarity measure. To prevent combinatorial explosion, only the top 1000 correspondences are kept after the similarity measure threshold is applied.

After filtering by similarity measure, most of the correspondences left should be on the measured ground. These remaining correspondences are used to detect the ground plane. A histogram is generated using the z-component of each of the 3D scene points that matched the ground plane. Typically, over 90% of correspondences fall on the measured ground, resulting in a histogram with a large peak indicating the elevation of the ground. A z-direction histogram is created using a moving window that scans the available height range. The window size is fixed at 2 meters and the bin spacing is set at 0.2 meters. The presence of a potential peak is detected when more than 25% of all scene points fall in the window. Since the peak should follow shortly, the bin spacing is reduced to 0.1 meters in order to refine the location of the peak window. Given that the peak window found contains more than 70% of all the correspondences, ground detection is considered to be successful.

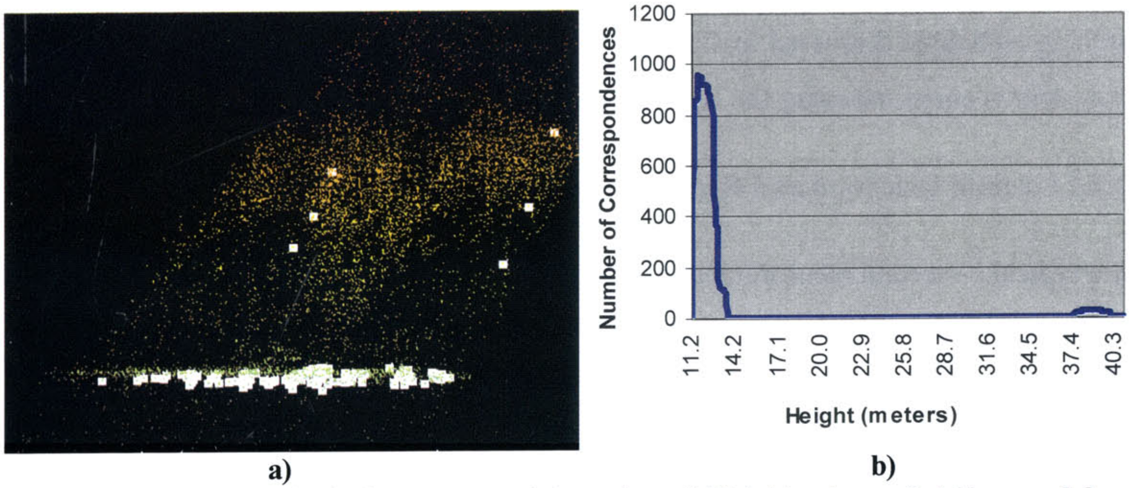


Figure 3-5: Example of spin-image ground detection. a) Height color-coded ¼-second frame of a ground target under canopy cover shown along the z-direction. The green to red pixels represent the measured data, while the larger white pixels represent the point correspondences found after spin-image matching the scene data to the plane model.
b) The resulting correspondence z-histogram of the ¼-second frame shows a single peak, which corresponds to the ground elevation; no peak exists for the tree cover since the spin-image matching algorithm correctly filtered out those measurements.

Figure 3-5 shows an example of spin-image based ground detection for a ¼-second frame, along with the resulting correspondence z-histogram. From the z-histogram, one can see that most of the correspondences fell on the measured ground; very few correspondences fell on the trees, even though they account for up to 90% of the data in the scene. Thus, the spin-image algorithm was able to successfully filter out most of the clutter due to the canopy cover and correctly detect the ground.

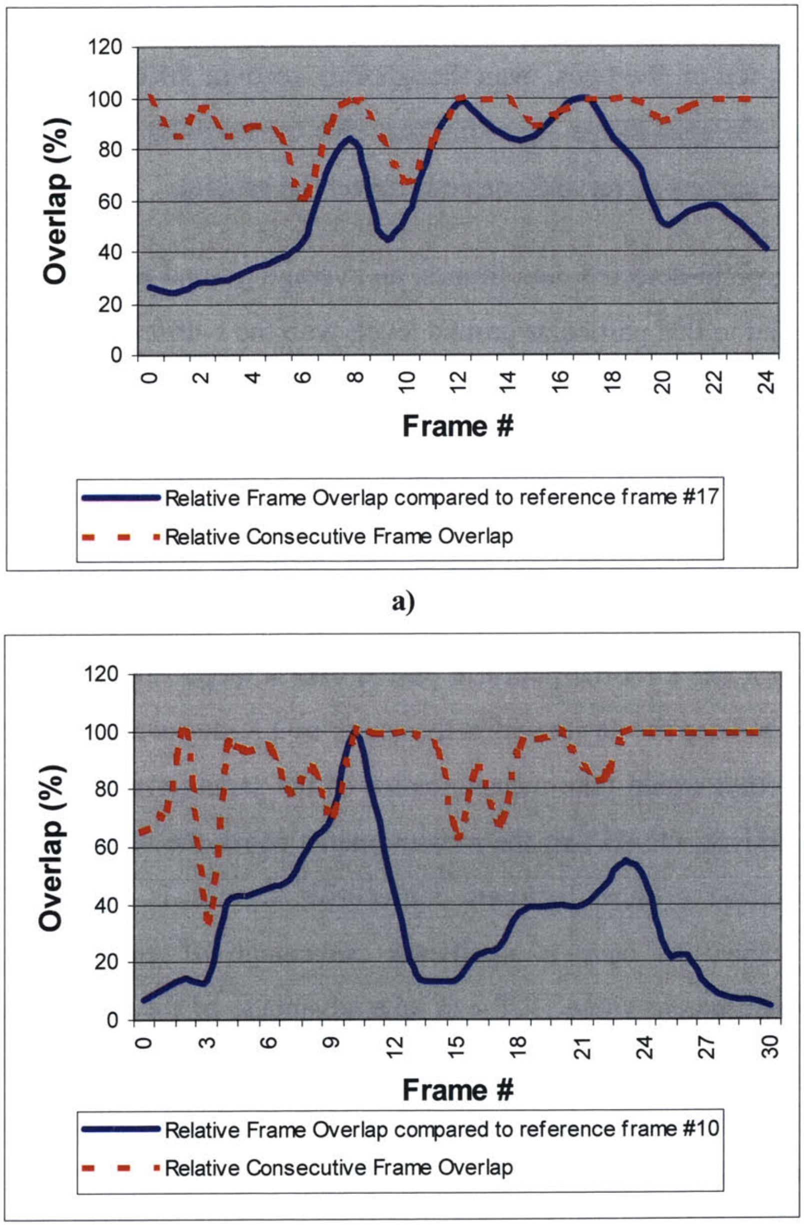
Given a set of ground-detected data frames, an average ground level is found and all the frames are aligned to that particular ground level; with the z-drift error corrected, only the xy drift error remains.

3.2.2.2 Variable-Overlap ICP Registration

The ICP algorithm described in 2.3.3 can be used to correct xy-drift errors. A registration algorithm was developed based on our method for data collection (See Section 3.1.2) As described in 3.1.2, the airborne platform passes over a target and constantly adjusts the gimble position to keep the target within the center of a Risley scanning pattern. Thus, an ideal data collection would take measurements of the same ground region from several different perspectives. Given that the measurements cover the same ground region, and that the perspective change between consecutive frames is relatively small (5-10 degrees), we expect to have a significant percentage of ground overlap between consecutive measurement views. ICP can take advantage of the high ground overlap to correct the xy-drift errors present in between the data frames. The procedure to align two data frames is described below in pseudo-code:

```
Align (data_frame N, data_frame M)
    1. Crop N and M in height at a range of 1<z<5m above the computed ground level.
    2. Compute the number of points left in N and M. Set the number of possible point
       matches to the maximum of the number of points left in N and M.
    3. Run ICP with a Search Voxel Size of 2.0 meters (this allows a point in N to match to a
       point in M if the points are at most 1.0 meter apart in any of the 3 Cartesian directions.)
    4. Compute the overlap percentage between data set N and data set M as the final number
       of ICP point-matches found divided by the maximum number of possible point-matches.
```

For measurements of targets out in the open, the relative ground overlap between consecutives frames should be very high, close to 100%. For targets underneath heavy canopy and/or camouflage cover, the consecutive frame ground overlap can vary from approximately 50% to 100%. (see Figure 3-6, dashed line).



a)

b)

Figure 3-6: Comparison of Relative Frame Overlap. a) The figure shows the relative frame overlap for a target in the clear. The dashed curve plots the consecutive frame overlap while the solid curve plots the overlap of a frame to reference frame # 17. b) The figure shows the relative frame overlap for a target underneath heavy canopy cover. The dashed curve plots the consecutive frame overlap while the solid curve plots the overlap of a frame to reference frame # 10.

As the change in perspective between two ¼-second data frames increases, the two respective frames are less likely to have areas of surface overlap due to varying canopy occlusion and change in viewing perspective. Over the optimal JIGSAW angular diversity of ±30 degrees to nadir, the ground overlap between a particular reference frame and the remaining frames drops sharply from close to 100% to less than 5% as the change in perspective increases (see Figure 3-6, solid lines). Figure 3-6 above shows a comparison between the relative overlap of consecutive frames versus the relative frame overlap to a particular reference frame. The comparison is shown for a target in the clear and a target underneath heavy canopy cover.

The general trend drawn from Figure 3-6 is as expected: the larger the change in viewing angles between two frames, the smaller the relative overlap. Given the overlap versus viewing angle statistics, there are two approaches to be considered for registration: consecutive frame registration and reference-frame registration. The advantage of consecutive frame registration is that the relatively high overlap between consecutive data frames should allow ICP to easily align the data frames. However, one major disadvantage of this registration approach is that each frame N is only well aligned compared to the next frame, N+1; no guarantee exists that frame N is optimally aligned to frame N+2, N+3, and so on. This disadvantage is particularly severe for our data sets since the registration error is present almost exclusively along one direction, namely the flight direction. Thus, even though ICP might remove a large portion of the drift between consecutive frames, some drift will still remain. Since the drift is in a constant direction, the error will tend to accumulate over several frames, leading to large registration errors.

However, the directional drift is not a problem for reference-frame registration: since the frames are aligned against a single frame, the resulting registration should not be affected by the drift direction. Thus, even though the "reference-frame to frame" overlap drops sharply over an angular diversity of ±30 degrees as compared to the consecutive frame overlap, the registration of the entire data should be considerably better. To test our hypothesis that reference-frame registration is better suited to correct our characteristic one-directional drift, we implemented preliminary versions of both registration approaches. Our initial results, shown in Figure 3-7, support the above conclusions.

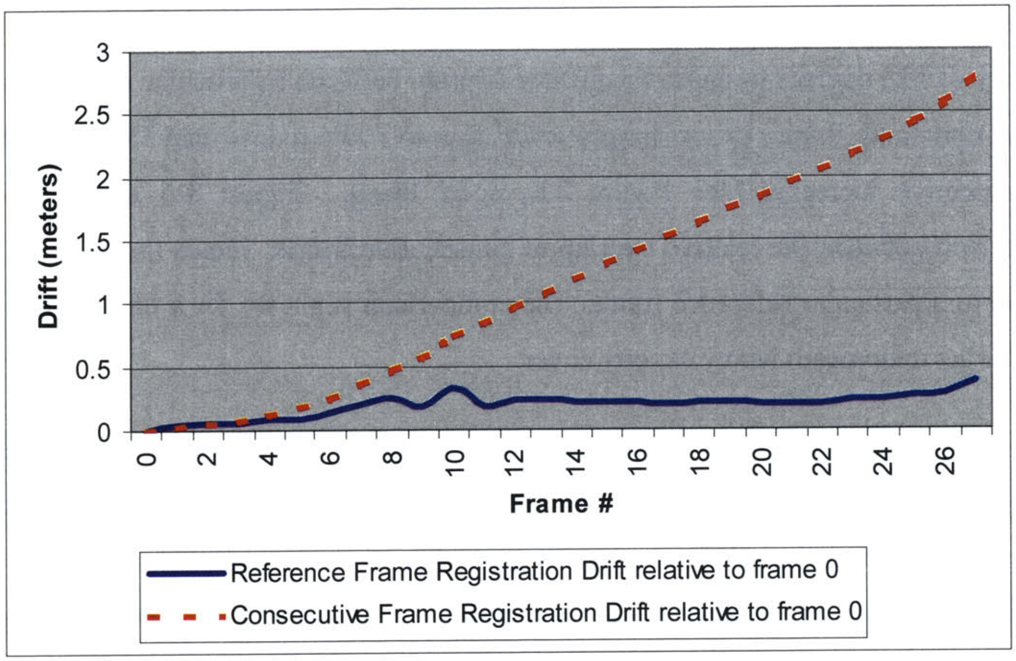


Figure 3-7: Registration Performance Comparison. Remaining drift after registration for data collection C20-F01-P01: the dotted curve represents the remaining drift after consecutive frame registration relative to frame 0; the solid curve is the remaining drift after reference frame registration relative to frame 0.

Thus, the implemented registration algorithm is based on reference-frame registration. For this approach, a reference frame is automatically selected from the group of frames to be registered. The reference frame is chosen based on an estimate of the target coverage in each frame. The frame with the highest target coverage is selected as the reference frame and the remaining frames are registered to that particular reference frame. Target coverage is used to select the reference frame because high target coverage increases the likelihood of surface overlap between the reference frame and the rest of the frames.

As mentioned previously, the primary purpose of registration is to reconstruct the ground target. Thus, we are interested in registering data right above the ground level. Since the ground in most of the measurements is flat and does not have a distinct spatial structure that ICP can lock onto, it is removed from each data frame. Removing the ground plane has several benefits for ICP registration, namely a performance speedup and a smoothing effect in pose-distance space that reduces the number of local minima, thus increasing the likelihood of ICP finding the optimal global minimum. Structures high above the ground level, such as canopy cover are also removed before ICP registration: canopy measurements could be very noisy due to wind conditions and variations in viewing perspectives. Based on the height of our target models, only measurements located in a height range of 1 to 5 meters above the measured ground plane are considered for ICP registration. The resulting height-cropped data frames are used to estimate target coverage for the reference frame registration approach. The number of points remaining in each frame after height-level cropping is used as an estimate of target coverage; the frame with the most points is declared the reference frame.

The registration process flow diagram is shown in Figure 3-8. First, all the frames are sub-sampled using 10-20cm cubic voxels in order to reduce the processing time. The reference frame R is then automatically selected from among the given data frames as specified above. Next, the frames closest in viewing angle to the reference frame are registered (i.e. frame R-1 and frame R+1). The registration is spread outwards from the reference frame, to frame R-2, R+2, and so on. As the registration spreads outwards, the transformations of the previous registered frames are applied: thus, before ICP-registering frame R-2 to frame R, the transformation found for R-1 to R is applied. By applying the previous transformations, only the drift between R-2 and R-1 now needs to be corrected by ICP.

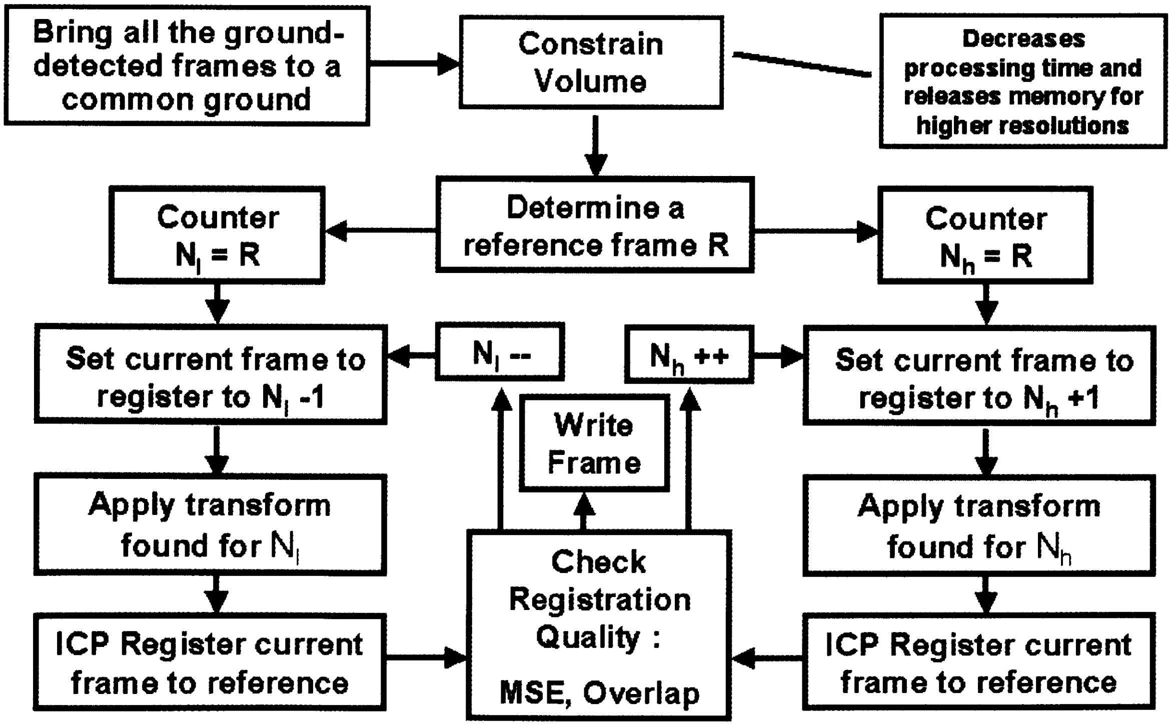

Bring all the ground-detected frames to a common ground
Constrain Volume
Decreases processing time and releases memory for higher resolutions
Counter N_l = R
Determine a reference frame R
Counter N_h = R
Set current frame to register to N_l -1
N_l --
N_h ++
Set current frame to register to N_h +1
Write Frame
Apply transform found for N_l
Check Registration Quality : MSE, Overlap
Apply transform found for N_h
ICP Register current frame to reference
ICP Register current frame to reference

Figure 3-8: Reference-Frame Registration Process Flow Diagram

A frame N is considered registered to the reference frame R if it meets certain overlap and MSE criteria. The minimum overlap criterion for the registration to be considered valid is that the current frame must have 15% overlap with the reference frame. The MSE criterion is based on the Cartesian distance between matching points in frame N and frame R. Since MSE is a function of the resolution of the data sets, the MSE threshold is set to 1 times the data resolution.

3.2.3 Surface Reconstruction Algorithm

Given a registered data set of 3D points, we need a surface-based representation of the data as input to the spin-image recognition algorithm. In our representation, surface shape is described by a collection of oriented points, 3-D points with associated surface normals.

The surface reconstruction algorithm is as follows:

1. The registered data set is sub-sampled using 10cm cubic voxels in order to have a uniform spatial density as required by the spin-image algorithm. Points falling into each voxel are averaged to create a single average 3D point.
2. Using the sub-sampled data set, compute an estimate of the surface resolution by finding the distance to the closest 8th neighbor of each point and taking the median distance value of the data set. The surface resolution estimates the mesh resolution variable used by A. Johnson.
3. A local surface normal is computed for each 3D point using the local point neighborhood. The local point neighborhood includes points that are at most 2 times the resolution distance away from the respective 3D point. The local point neighborhood distance is based on the data resolution in order to assure that most of the points will have enough neighbors to compute a local surface. The local surface normal is computed by applying the Singular Value Decomposition algorithm to the local points.

The result of the surface reconstruction algorithm is an oriented 3D point data set that has uniform spatial density. The 3D oriented data set can be input into the spin-image algorithm to perform target detection and recognition.

3.3 Data Preprocessing Results & Discussion

3.3.1 MPRCP Results

Typically, the raw sensor data, when transformed into Earth-fixed coordinates, forms a solid data cube. Figure 3-9-a shows a height color-coded, ¼-second frame of data recorded at Huntsville in January 2003. The raw data set is a measurement of a M-60 underneath heavy canopy cover. Without any data filtering, a human viewer cannot readily detect the structure of the trees, ground and more importantly, the M60 tank. Figure 3-9-b shows the same data set after applying the MPRCP range-filtering algorithm. Visual comparison of the two figures shows a large improvement in quality: the canopy structure, tree branches and trunks as well as target structure becomes more readily apparent.

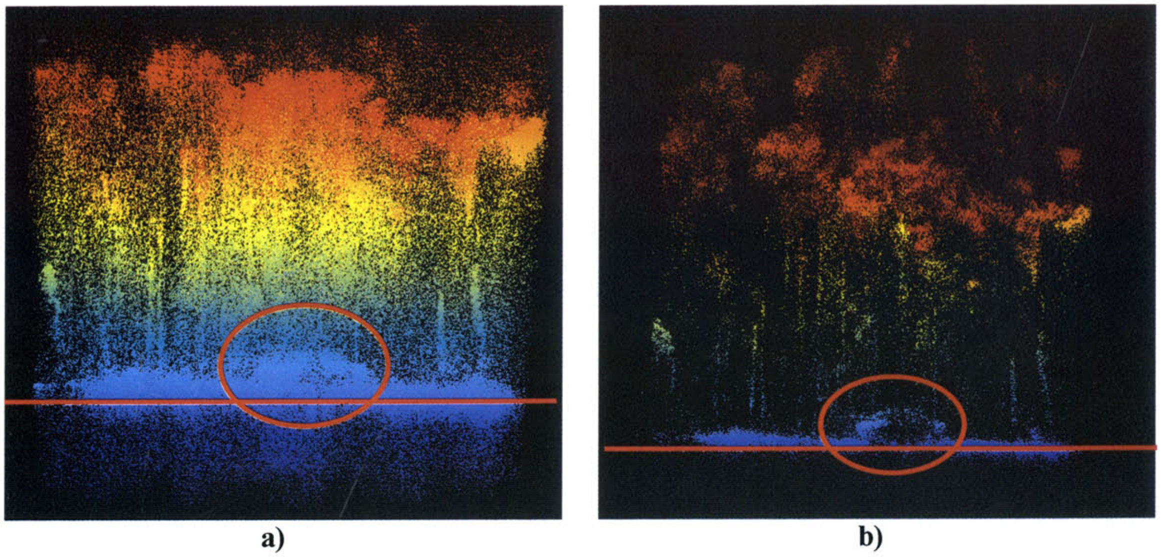


Figure 3-9: Example of MPRCP Results. a) Height color-coded raw data frame shown along the z direction; the red line represents the ground level, while the red oval indicates the target location, b) Height color-coded MPRCP processed frame shown along the z-direction. Compared to the raw data, the MPRCP data has most of the range-noise removed.

3.3.2 Registration Results

After MPRCP processing each ¼-second data frame, the entire data pass is registered according to the registration algorithm in Section 3.2.2. Figure 3-10 below shows 50 unregistered ¼-second data frames of an M2-A3 APC under canopy cover. The data is height color coded and displayed along the z-direction to indicate ground drift. Over the time of the data collection, the ground drifts by approximately 5 meters in the height direction.

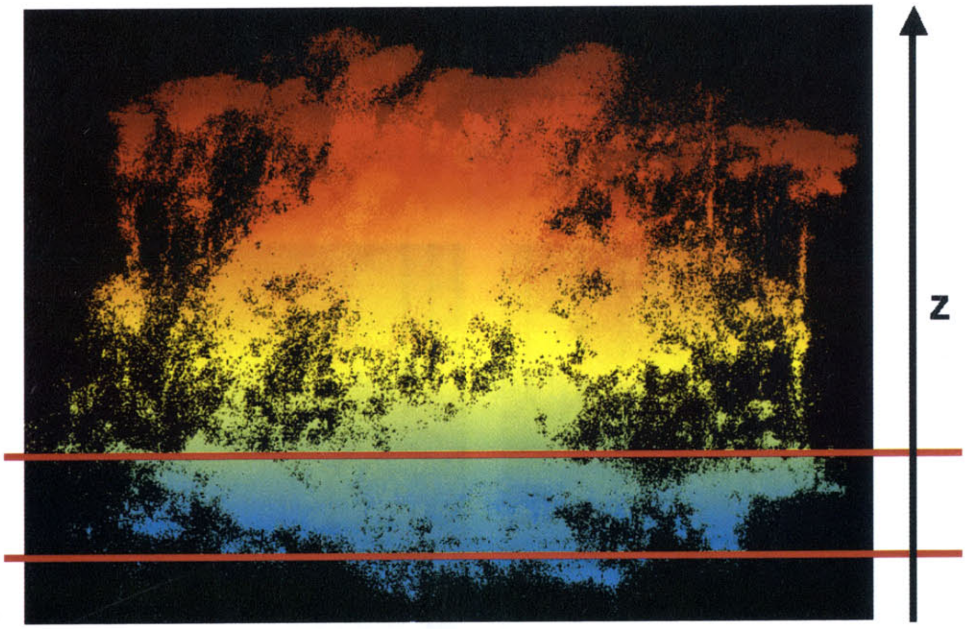
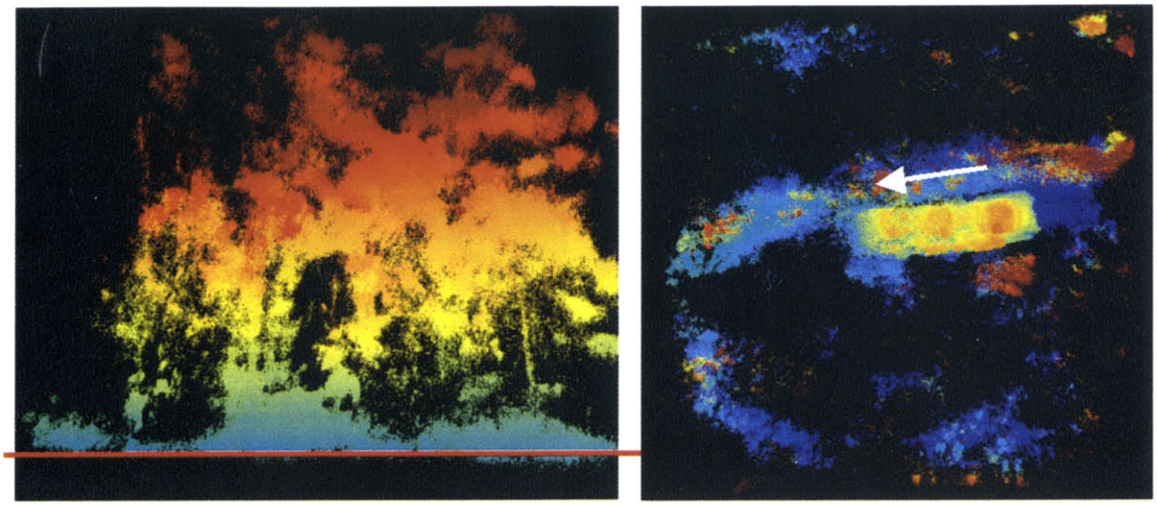


Figure 3-10: Height-color coded unregistered MRPCP processed data consisting of 50 ¼-second frames. The data is displayed along the z-direction to emphasize the drift of the ground plane. The two red lines indicate the range of ground plane elevations, which corresponds to 5 meters of drift in the z-direction.

As described in Section 3.2.2, our registration algorithm first corrects errors in the z direction and then in the xy direction. Figure 3-11 shows the data pass after z-registration. In Figure 3-11-a, the ground level is very thin in the z-direction, as most of the z-drift has been removed. However, the data, as shown from a bird's eye view in Figure 3-11-b, still has large drifts in the xy direction resulting in severe streaking of the measured M2-A3 APC.



a) b)

Figure 3-11: Height-color coded MPRCP processed data after z-registration. a) A view of the data along the z direction to show the successful registration of the ground; the ground level is indicated by the horizontal red line. b) An orthographic view of the data, with the trees cropped out. The white arrow indicates the xy registration drift that remains to be corrected.

Figure 3-12 shows the data pass after applying reference-frame ICP registration. The figure shows the data pass from several viewing perspectives to assess the goodness of the overall registration. From the collection of viewing perspectives, it becomes apparent that the target has been correctly registered: the target's edges are very sharp and detailed structure of the APC such as the turret, missile launcher, hatch, and front lights can easily be discerned. An M2-A3 CAD model is shown in Figure 3-12-c to provide a visual comparison to the measured target. The measured target's structure and relative dimensions match the M2-A3 truth model, confirming that the registration was successful.

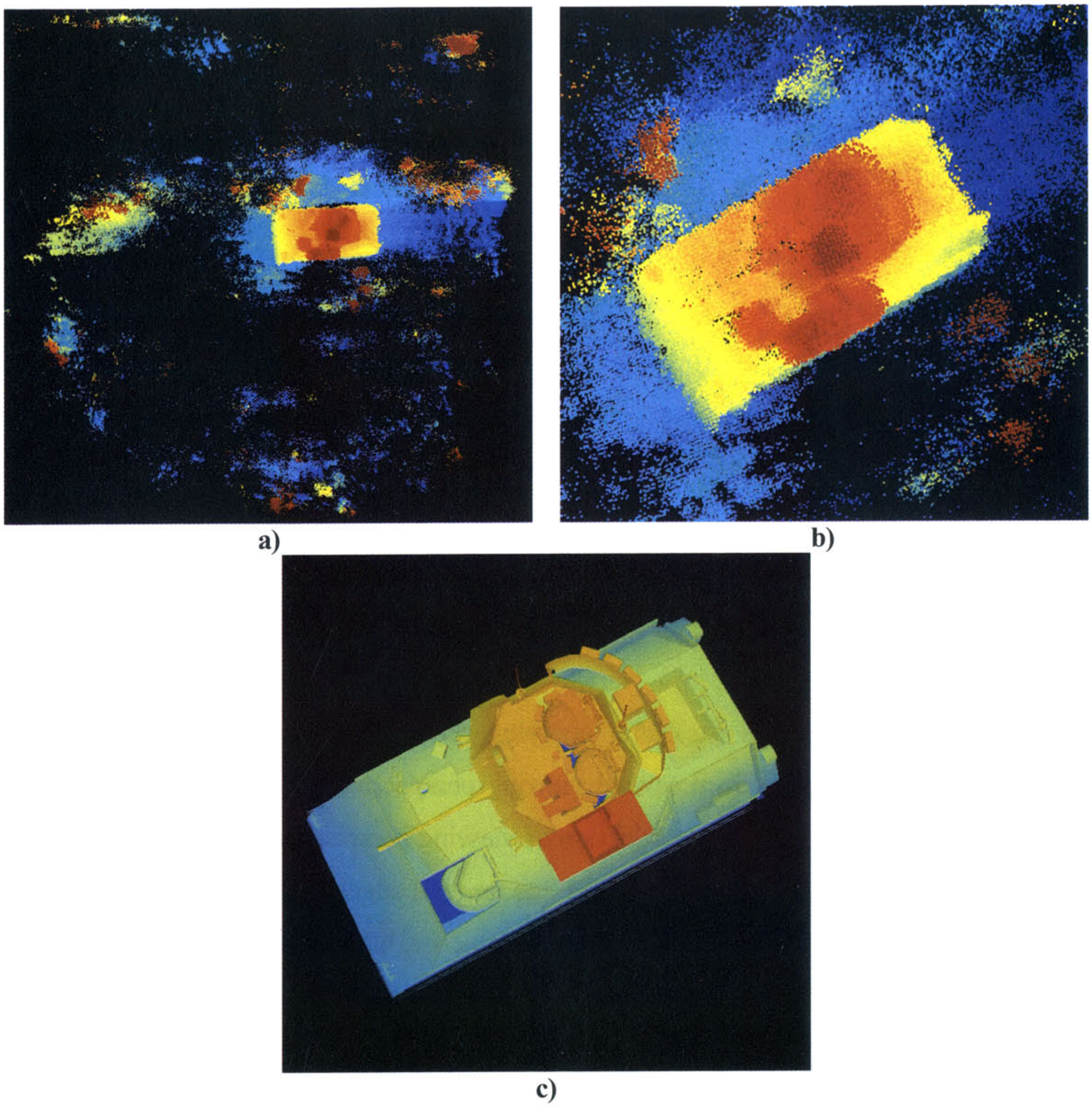


Figure 3-12: Height color-coded orthographic perspective of the fully registered data pass. a) View of the entire data pass, with the trees cropped. The measurement of the M2-A3 APC is shown in shades of yellow to red. b) A close-up look of the M2-A3 APC measurement. The APC's structure, such as the body, turret, missile launcher, hatch, and front lights, can easily be discerned. c) An M2-A3 CAD model to provide a visual comparison to the measured M2A3. The measured target's structure and relative dimensions match the M2-A3 truth model, confirming correct registration.

In order to accurately convey the overall quality and time performance of the implemented registration algorithm, we have provided a table that summarizes the results for 53 registered data collections. The registration algorithm has several parameters that can be adjusted. All other parameters are set according to Sections 2.3.1 and 3.2.2. The adjustable parameters, along with their default values, are as follows:

1. Spin-Image resolution – indicates the data sub-sampling resolution used to perform ground detection / registration (Set at 1.0 meters).
2. Maximum number or ICP iterations. (Set to 50 iterations).
3. ICP resolution – indicates the data sub-sampling resolution used to perform ICP registration (Set to 0.1 meters).
4. Maximum MSE – MSE threshold value for considering a particular frame to be correctly registered to the reference frame (Set to $0.10 \mathrm{m}^2$).
5. Minimum Overlap – Minimum overlap percentage required between a particular frame and the reference frame for the registration to be considered valid. (Set to 15%)

Table 3-1 summarizes the registration results. The registration algorithm was run on a Intel Pentium IV Xeon 2.0 Ghz machine. The average time taken for each registration stage is given as a multiple of real time, where real time is defined as the time to collect the data set. R_t, the registration time versus real time, is defined as follows:

$$R_t = \frac{\text{Time to Register Data}}{\text{Time to Collect Data}} \qquad \textbf{(3.2)}$$

Quality performance is indicated by the fraction of data that ICP-registered along with average ICP MSE found for the registration. The fraction of data collection registered is defined as the number of registered ¼-second frames divided by the total number of ¼-second frames.

	Average Ground Registration Time (versus real time)	Fraction of Data that Ground Registered	Average ICP Registration Time (versus real time)	Fraction of Data that ICP Registered	Average MSE (m^2)
Demo 1& 2 MPRCP Data	26.2X	0.90	7.5X	0.59	0.051

Table 3-1: Registration Time and Quality Performance for 53 data sets collected using the JIGSAW sensor.

The registration time results in Table 3-1 indicate a total registration time of approximately 34X Real Time. Thus for a 10-second data collection, registration would take about 340 seconds to complete. Based on the ICP registration constraints (Maximum MSE of 0.10m^2 and Minimum Overlap of 15%), approximately 59% of the data collection was successfully registered. The remaining 41% of the data collection that did not register was typically composed of frames at the beginning and end of the data set: these frames contain data measurements taken at angles larger than 30 degrees to Nadir, resulting in heavy canopy coverage and sparse target coverage. Due to sparse target coverage, these frames typically did not pass the minimum overlap requirement of 15%. Thus, even though the frames comprised 41% of the data collection, they typically had very low target coverage compared to the rest of the registered frames. As a result, their contribution to the registered target data is minimal. Another registration quality parameter is the Mean Squared Error found by ICP between the registered frames and the reference frames. The MSE value is indicative of the average drift that is left after registration. An MSE value of 0.051m^2 indicates an average expected drift of about 23 cm. Considering that the native data resolution is about 10 to 20 cm, the remaining drift is relatively small. The significant fraction of data registered coupled with a small MSE value demonstrates that the registration process was successful.

3.3.3 Surface Reconstruction Results

Figure 3-13 below shows an example of surface reconstruction for a measured data set of an M60 tank. From the figure, we can visually discern that the computed surface orientation is a good estimate of the target's local surface.

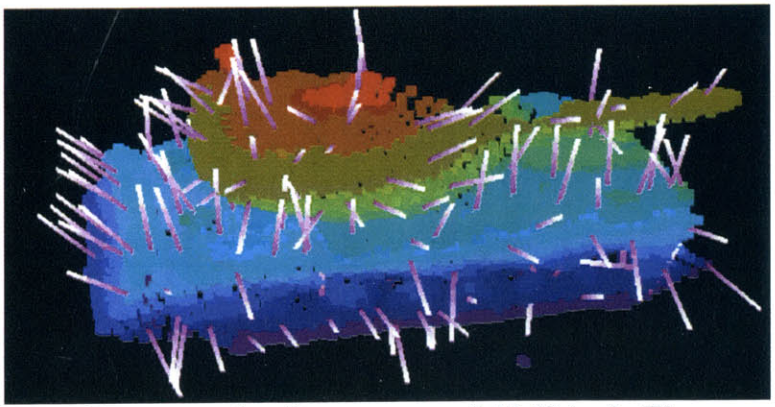


Figure 3-13: Surface Reconstruction for a measured M-60 tank. The scene data is height-color coded in shades of deep blue through green and red. Each purple line represents the surface normal found for a particular point. Only a small percentage of the surface normals are shown so that the underlying M60 measurement can still be visible.

Table 3-2 below summarizes the quality and timing performance of the surface reconstruction algorithm. The estimated surface normals are within 15 degrees compared to the true surface normals obtained from the target CAD models (see Chapter 4). As discussed in Chapter 2, a spin-image is created based on a corresponding oriented point basis. Thus, the accuracy of the surface normals affects spin-image creation and subsequent spin-image correlation between the scene and the model points. Johnson's study of the effects of scene noise and surface normal error on spin-image correlation has shown that the errors in surface normal orientation should not have a great effect on spin-image correlation [24]. Based on Johnson's results, summarized in Figure 9-17 of his PhD thesis, an error in surface normal of approximately 15 degrees will result in the normalized correlation value to decrease by no more than 2% as compared to the ideal case where no surface normal error exists. Thus, our surface reconstruction provides a good estimate of surface orientation that can be effectively used by the spin-image recognition algorithm.

	Scene Description	Scene Points	Scene points sub-sampled @ 10 cm	Average Error In Estimated Normal Direction (Degrees)	Time to Subsample (seconds)	Time to Compute Normals (seconds)	Total Time (seconds)	Total Time versus Real Time
Field Data	BMP-1 scene (c5-f10-p3)	42605	11936	14.3	0.45	14.17	14.62	2.34
	BTR-70 (c5-f10-p4)	38551	13514	15.7	0.47	17.67	18.14	2.90
	HMMW	24219	3966	13.9	0.39	2.23	2.62	N.A.
	M1-A1 Eglin	13120	3645	16.4	0.11	1.17	1.28	N.A.
	M2-A3 scene (c8-f4-p20)	14556 8	19858	15.8	1.41	37.26	38.67	3.09
	M-35 (C5-F10-P5	34699	10166	13.4	0.297	11.83	12.127	1.94
	M60 field tank 1	5589	3932	15.7	0.09	1.08	1.17	N.A.
	M60 field tank 2	4045	2982	17.4	0.09	0.78	0.87	N.A.
	t-72 (C5-F19-P3)	36543	36543	14.1	0.56	36.52	37.08	4.01

Table 3-2: Surface Reconstruction Timing Performance.

In summary, we have successfully developed a processing algorithm to reconstruct measured scenes from multiple viewing perspectives. In addition, we developed a surface reconstruction procedure in order to generate a 3D oriented data set that can now be passed to the spin-image detection and recognition algorithms in order to identify the scene target(s) from among a library of target models.

In the next chapter, we will discuss the modeling procedure used to generate 3D oriented point data sets from CAD models of target objects. Based on a visual analysis of clutter and occlusion in our particular measurement scenes, an optimal set of spin-image generation parameters is selected in order to maximize recognition performance.

Chapter 4: Object Modeling for Generation of Spin-Image Libraries

The JIGSAW program has approximately 10 targets of interest, ranging from trucks and APCs to tanks and missile launchers. The CAD models of the specific targets are shown below in Figure 4-1. The model library contains two large target classes, namely APCs and tanks. The APC target class is composed of the BMP-1, BMP-2, BTR-70 and M2 vehicles. The tank class includes the M1A1, M60 and T72.

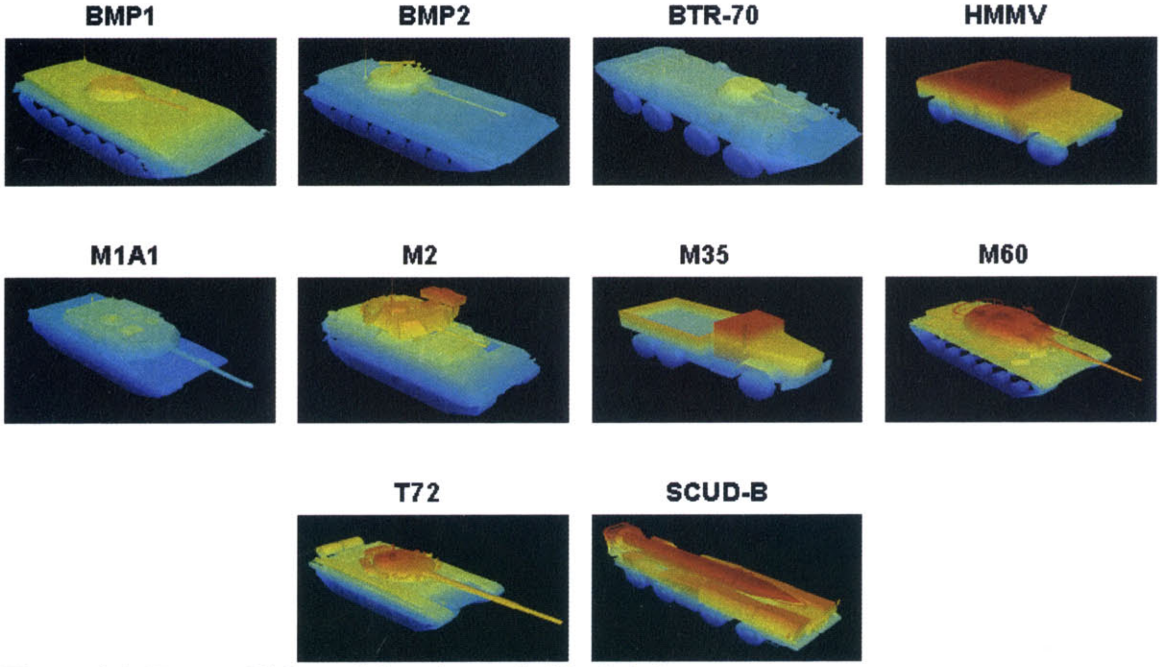


Figure 4-1: Target CAD models color-coded by height.

Based on the above CAD models, a target model library was constructed to simulate an ideal 3D LIDAR signature of each target. The simulated targets are then represented in the spin image representation as 3D oriented points with associated spin-images. The resulting model spin-image library is used to compare the models to measured scenes in order to recognize and identify the scene target.

4.1 Modeling Procedure

We implemented an object modeling procedure to reconstruct the surface of a target as seen from multiple views. We created an openGL application that recorded the surface measurement of the target CAD model from multiple views by capturing the screen z-buffer. For each view, the resulting z-buffer returned the absolute xyz position for each rendered pixel on the CAD model surface. Each xyz position returned by the z-buffer also has an associated normal, based on the surface normal of the respective triangular CAD model element that generated that particular pixel. In order to associate a normal to a particular pixel, the index of each of the triangular elements of CAD model was used to generate a unique RBGA 32-bit color value. The RGBA color value of each pixel was then read from the z-buffer in order to decode the index of the particular triangular element and find its associated surface normal. Simulated measurements were taken from 9 positions, namely: from nadir, looking straight down on the target, and 8 views at yaw angles spaced 45 degrees apart, at a pitch of 30 degrees. (see Figure 4-2)

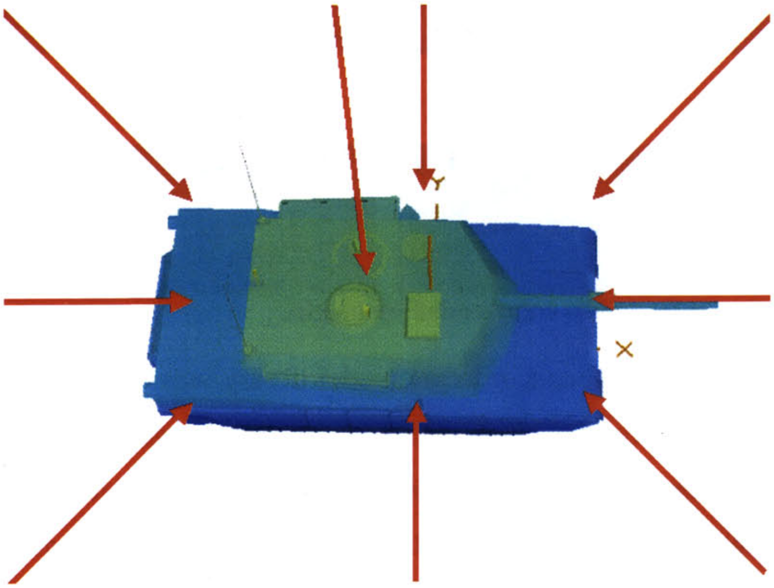


Figure 4-2: Viewing directions used to simulate 3D Ladar data of an object from a CAD model.

The result of the modeling procedure is a high-resolution 3D oriented point set that captures most of the viewable surface of a target that could be seen by an airborne LIDAR sensor. The 3D oriented data set is then sub-sampled using 10cm or 20cm cubic voxels in order to have a uniform spatial density as required by the spin-image algorithm. Based on the sub-sampled data set, a spin-image stack is created.

The above procedure is applied to each JIGSAW target and a spin-image library is created. The spin image stacks for each target are generated using the same parameters for locality and resolution (i.e. spin-image width, support angle and bin size).

4.2 Results

Table 4-1 summarizes the resulting model data sets obtained from the 3D simulation. The estimated surface resolution parameter is computed based on the method described in Section 3.2.3 on scene surface reconstruction.

		Number of Points in the Model Data Set									
Sub-sampling Voxel Size (m)	Estimated Surface Resolution (m)	BMP1	BMP2	BTR-70	HMMV	M1A1	M2	M35	M60	SCUD-B	T72
0.1	0.125	10366	9692	11444	5035	16394	14669	10146	18778	23916	13454
0.2	0.25	2239	2056	2453	1255	3761	3223	2368	4035	5213	2842

Table 4-1: Resulting 3D oriented point data sets for the given target models for two sub-sampling voxel sizes.

Based on the above 3D oriented point data sets, several spin-image models libraries are generated for a range of spin-image parameter values. Table 4-2 shows the resulting spin-image libraries for a variety of spin-image parameter combinations. In the table, several spin-image generation parameters are considered, namely data resolution, spin-image width, bin-size, and support angle. Based on each set of spin-image parameters, a spin-image model library is computed. The spin-image library statistics are described by the *Average Fraction of Model Surface Coverage* captured by the spin-image, the *Average Lambda Spin-image Fill Factor* and the *Average Stack Creation Time*. Based on these resulting statistics, optimal sets of spin-image generation parameters are selected in order to maximize detection and recognition performance while meeting our recognition run-time goals.

Average number of points per model	Data Resolution (meters)	Spin Image Width (pixels)	Spin Image Bin-Size (meters)	Spin Image Distance of support (meters)	Angle of Support (degrees)	Average Fraction of Model Surface Coverage	Avg. Lambda (Spin-Image Fill Factor)	Avg. Spin-image Stack Create Time per model (secs)
13390	0.1	5	0.25	1.25	90	0.09	11.05	20.77
13390	0.1	5	0.25	1.25	135	0.12	11.2	23.47
13390	0.1	5	0.25	1.25	180	0.13	11.25	21.26
13390	0.1	10	0.25	2.5	90	0.24	38.25	53.78
13390	0.1	10	0.25	2.5	135	0.35	39.25	64.60
13390	0.1	10	0.25	2.5	180	0.38	39.5	59.58
13390	0.1	15	0.25	3.75	90	0.37	70.8	76.29
13390	0.1	15	0.25	3.75	135	0.57	72.9	95.58
13390	0.1	15	0.25	3.75	180	0.63	73.2	91.45
13390	0.1	20	0.25	5	90	0.45	104.35	92.36
13390	0.1	20	0.25	5	135	0.71	107.85	117.81
13390	0.1	20	0.25	5	180	0.81	108.3	115.29
2945	0.2	5	0.5	2.5	90	0.26	11	1.537
2945	0.2	5	0.5	2.5	135	0.39	11.1	2.138
2945	0.2	5	0.5	2.5	180	0.43	11.1	2.244
2945	0.2	10	0.5	5	90	0.45	32.35	2.389
2945	0.2	10	0.5	5	135	0.72	33.1	3.636
2945	0.2	10	0.5	5	180	0.83	33.25	4.076

Table 4-2: Spin-Image Generation Statistics and Timing for the JIGSAW model library for data resolutions of 10cm and 20cm. The table rows shaded in gray represent some of the optimal parameter combinations that are to be used for the detection and recognition model libraries.

4.3 Discussion

Based on the results shown in Table 4-2 and a visual analysis of clutter and occlusion in our particular measurement scenes, the default library used for target recognition had a data resolution of 10cm, bin size of 25cm, support distance of 2.5 meters, and a support angle of 90-degrees. For target detection, the optimal spin-image library had a data resolution of 20cm, bin size of 50 cm, support distance of 2.5 meters and a support angle of 90-degrees.

The recognition data resolution was determined based on the native sensor resolution, the required run-time performance and the overall target dimensions. Given a nominal sensor range resolution of approximately 7.5 cm, a recognition time goal of 10 minutes and target dimensions with footprints in between 2x4 meters to 3x12 meters, the data resolution for recognition was set to 10cm. A 10cm resolution should provide enough detail on the target models to allow correct target identification, while achieving the required timing performance set out in the goals. For target detection, the scenes can be up to 2 orders of magnitude larger than target recognition scenes. In order to achieve a reasonable detection time, the detection data resolution must be lowered. Based on our detection results, (presented in Chapter 6), a detection data resolution of 20cm provides enough detail to correctly detect the target while achieving reasonable detection times on the order of 1-2 minutes per model search.

According to A. Johnson, the bin size should be a factor of the estimated surface resolution; a typical bin-size is 2 times the estimated surface resolution. Thus the bin-size for target recognition was set to 25cm, while the bin-size for target detection was set to 50 cm.

The support distance is based on the physical size of the target models. Typically, the support distance is set on the order of the model. Shorter support distances are recommended for heavily cluttered scenes. Since targets in typical JIGSAW 3D scenes are under dense canopy cover, surrounded by trees, shrubs and sometimes veiled in camouflage nets, we expect to have a heavy amount of clutter. Based on A. Johnson's spin-image matching clutter analysis model and our target dimensions, the support distance was set to 2.5 meters. The resulting spin images capture the local point density in a cylindrical volume with a 2.5-meter radius, and 2.5-meter height. Spin-images obtained from this volume should provide a reliable local description of the model. According to Table 4-2 the average percentage of model surface area captured by a spin-image with a support distance of 2.5 meters is 38%. (the percentage corresponds to a support angle of 180 for which all the points in the cylindrical volume are used to create a spin-image).

Thus, a significant amount of the spatial structure of the target models is captured in a single spin-image. Concurrently, the chosen support distance avoids spin-image matching degradation for heavily cluttered scenes. Given a heavily cluttered scene, only clutter within at most 2.5 meters from the target surface will be included in the spin image. Thus, the region of clutter that affects spin-image matching is well constrained around the target and should have a limited effect on the recognition of the object.

The same support distance is used for target detection and recognition because we expect to have the same amount of clutter around a target in a recognition scene as for a target in a detection scene. This should always be the case since the target detection algorithm processes large-scale scenes and finds regions of interest where a target is likely to be located; the ROI is then passed to the recognition algorithm. Therefore, the same data in the target region is utilized for both target detection and recognition.

The third spin-image parameter is support-angle. The support angle is used to lessen the effects of object self-occlusion due to a limited number of viewing perspectives. A nominal support angle can be determined based on the angular diversity of the target measurements. Typically, the support angle should be set as high as possible, within the range of 60 to 180 degrees. For JIGSAW measurements, the angular diversity ranges from 20 degrees to approximately 60 degrees. For a 0-degree angular diversity (i.e. single view), a nominal support angle is 60 degrees. For an angular diversity θ_{AD}, the support angle is computed according to the equation below:

$$\begin{gathered} \textit{Support Angle}(\theta_{AD}) = \textit{Support Angle}(0^o) + \theta_{AD} \\ \textit{where Support Angle}(0^o) = 60^o \end{gathered} \qquad \textbf{(4-1)}$$

The average angular diversity for JIGSAW measurements is approximately 30 degrees, leading to a nominal support angle of approximately 90 degrees. Thus, a support angle of 90 degrees is used for both target detection and target recognition spin-image model libraries.

In summary, we developed a Lidar simulation that creates a dense 3D oriented data set given a particular CAD model. We utilized the high-resolution oriented data set to create several spin-image model libraries for a range of spin-image generation parameters. Based on the results obtained from those particular spin-image libraries, we selected the optimal spin-image generation parameters to be used for the target detection and target recognition procedures. In the next chapter, we describe the implemented target recognition algorithm. We will then present and discuss the obtained target recognition results.

Chapter 5: Automatic Target Recognition

Given a region of interest within a large-scale scene, the ATR algorithm attempts to identify the target from among the targets in a model library or report a "none of the above" outcome.

5.1 Recognition Algorithm

The ATR system is based on the surface-matching algorithm described in Chapter 2. The algorithm takes a scene data set along with a spin-image model library. A scene spin-image stack is created using the same spin-image generation parameters as for the spin-image model library. A sub-sampling of the points is used to create the spin-image stack. The sampling ranges from 20% to 50% of all scene data points. The scene data points are not judiciously picked: the points are uniformly distributed across the given scene. Therefore, no feature extraction is performed to pick certain points. The scene spin-image stack is correlated to each spin-image model stack within the model library. The resulting correspondences are filtered and grouped according to Section 2.3.2. The resulting pose transformations are verified using the ICP algorithm and assigned a V_{GOF} value according to equation 2-11. The pose transformation with the largest V_{GOF} value is considered to be the final result of the scene to model comparison.

To quantify recognition performance, a probability of detection is defined for each model to scene correlation based on the V_{GOF} value of the best pose transformation. The probability of detection (Pd) that the scene ***s*** correctly matches model ***i*** in the model library ***mlib*** is defined as:

$$Pd(s,\, mlib_i) = \frac{V_{GOF}(s,\, mlib_i)}{\sum_{j=0}^{N} V_{GOF}(s,\, mlib_j)} \qquad \textbf{(5.1)}$$

For each scene to model library comparison, the probability of detection is split among the models and ranges from 0 to 1. For a given scene, the sum of the Pd values over all the models in the model library adds up to 1, unless a "none-of-the-above" outcome is reached. In the case of a "none-of-the-above" conclusion, the sum of the probability of detection values will equal zero, and each $Pd(s,\ mlib_i)$ will be defined to equal zero.

The higher the probability of detection for a scene to a model, the more likely it is that the model correctly matches the respective scene. Thus, the Pd value that falls on each model represents a confidence measure that the model matches the scene. For the purpose of quantifying recognition performance, we assign the model with the largest Pd value to be the final recognized target. Thus the recognition result for each scene ***s*** to model library ***mlib*** is defined as:

$$\mathrm{Re}\,cognized\,Model(s, mlib) = \max(Pd(s, mlib_i)) \qquad \textbf{(5.2)}$$

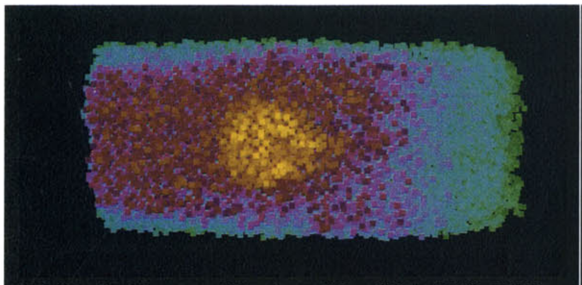
5.2 Results & Discussion

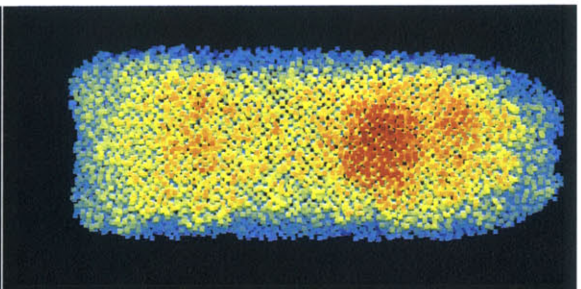
The ATR results are divided into two sections. The main section is devoted to the non-articulated ATR results obtained from the comparison of twelve measured data scenes to the target model library generated with the optimal spin-image parameters determined in Chapter 4. A second, smaller section will focus on the results of a limited study of articulated ATR.

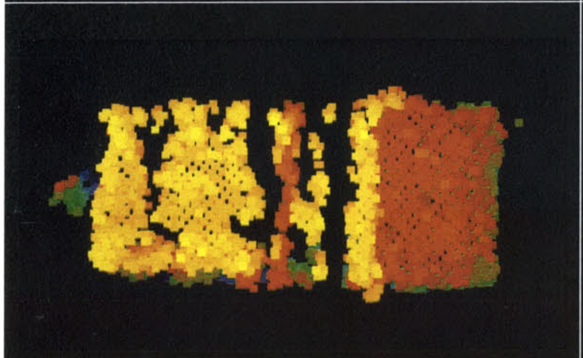
5.2.1 Non-Articulated ATR Study

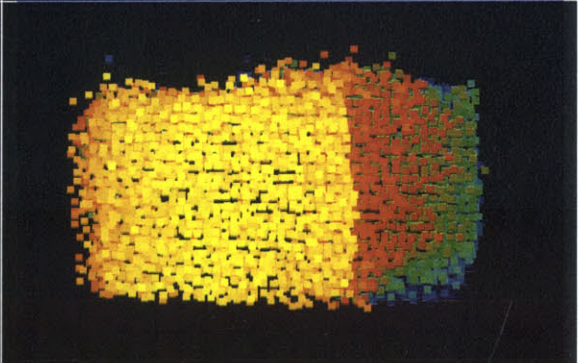
For the study of non-articulated ATR, we used the ten-object target model library presented in Chapter 4. Based on the spin-image generation results discussed in Chapter 4, the default library used for target recognition had a data resolution of 10cm, 25cm bin size, 2.5 meter support distance, and a 90-degree support angle.

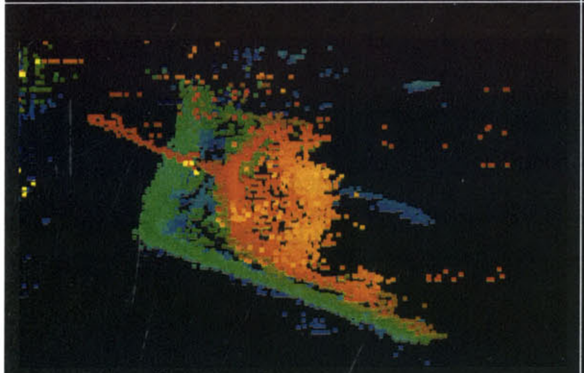
In order to determine the recognition performance, multiples scenes were analyzed. A probability of confusion matrix is utilized to show the recognition performance, wherein the confidence measurement Pd is shown on the main diagonal and errors on the off diagonals [36]. Twelve scenes were used to create the probability of detection confusion matrix. Target truth was known prior to data collection. Measured data for the following targets was used: BMP-1, BTR-70, HMMW, M1A1, M2A3, M-35, M60 and the T-72. Figure 5-1 below shows an orthographic projection of each of the twelve measured data sets, along with the target ID, date and location of measurement and campaign-flight-pass numbers.

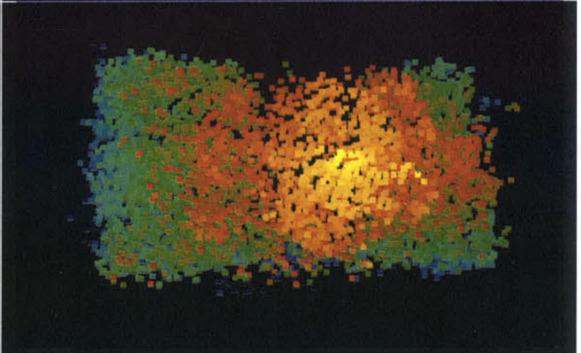

BMP-1, Huntsville Jan 03, C05-F10-P03

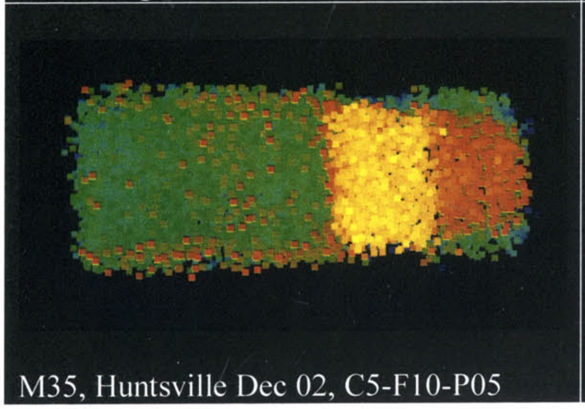

BTR-70, Huntsville Jan 03, C05-F10-P04

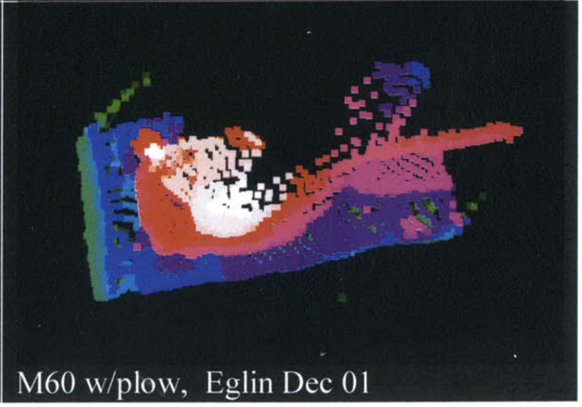

HMMV, RMF May 02

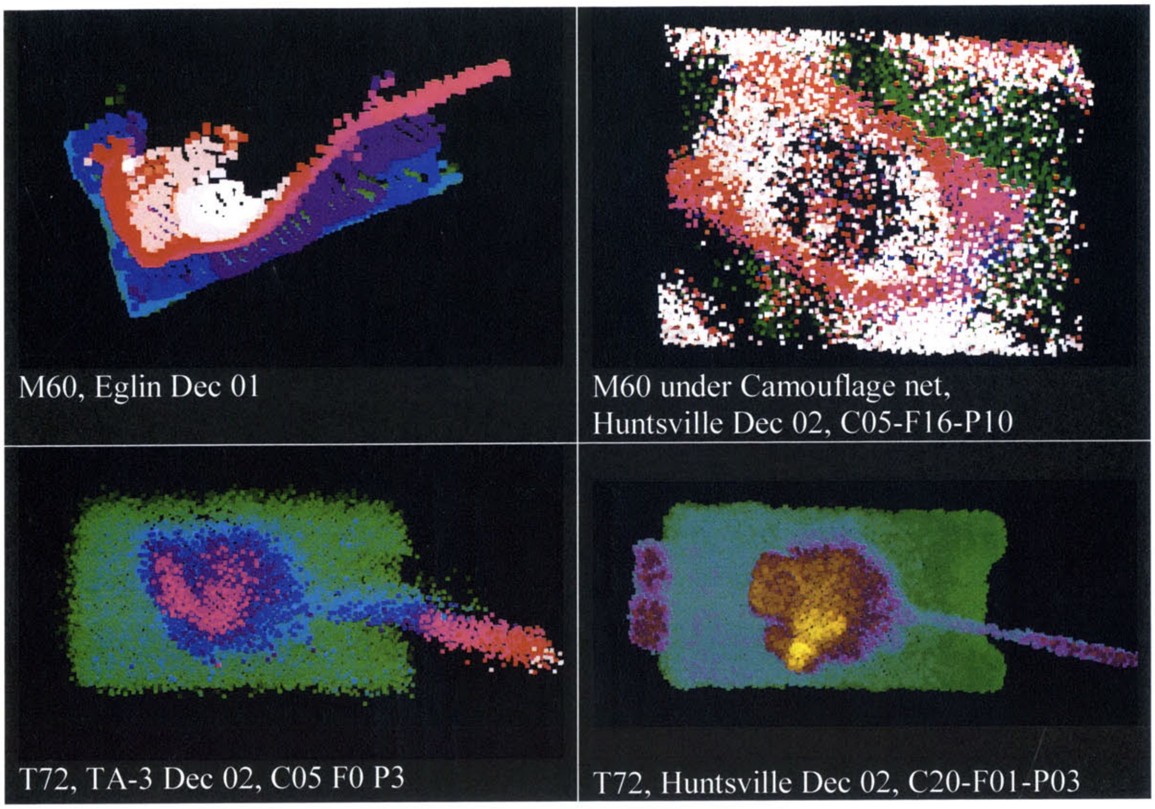

HMMV, Huntsville June-03, C8-F1-P10


M1A1, Eglin Dec 01


M2-A3, Huntsville Dec 02, C5-F10-P05


M35, Huntsville Dec 02, C5-F10-P05


M60 w/plow, Eglin Dec 01



M60, Eglin Dec 01

M60 under Camouflage net, Huntsville Dec 02, C05-F16-P10

T72, TA-3 Dec 02, C05 F0 P3

T72, Huntsville Dec 02, C20-F01-P03

Figure 5-1: Orthographic view of the twelve measured scenes with height color-coding.

Table 5-1 shows the probability of detection confusion matrix obtained from the comparison of the model library to each of the twelve scenes. Each row of the confusion matrix represents a scene to model library comparison; for instance the first row contains the comparison between a BMP-1 scene measurement and the model library.

				Models									
		Angular Diversity	Angular Views	BMP1	BMP2	BTR-70	HMMV	M1A1	M2	M35	M-60	SCUD-B	T72
Field Data	**BMP-1** C5- F10-P3	10°	16	0.61	0.39	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.02
	BTR-70 C5- F10-P4	10°	23	0.0	0.0	0.81	0.0	0.19	0.0	0.0	0.0	0.0	0.0
	HMMV RMF May 02	15°	4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
	HMMV C8-F1-P10	30°	10	0.0	0.01	0.0	0.92	0.0	0.0	0.0	0.0	0.0	0.07
	M1-A1 Eglin Dec 01	0°	1	0.0	0.0	0.0	0.0	0.83	0.0	0.0	0.0	0.0	0.17
	M2-A3 C5-F13-P07	20°	16	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
	M-35 C5-F10-P05	15°	12	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
	M60 w/plow Huntsville May02	0°	1	0.0	0.0	0.09	0.0	0.0	0.0	0.0	0.91	0.0	0.0
	M60 Huntsville May02	0°	1	0.0	0.03	0.0	0.0	0.0	0.01	0.0	0.96	0.0	0.0
	M60 C05-F16-P10	10°	12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.97	0.01	0.02
	T-72 TA-3 Dec 02	15°	105	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.00
	T72 C20-F01-P3	15°	29	0.0	0.0	0.0	0.0	0.13	0.0	0.0	0.0	0.0	0.87

Table 5-1: Probability of Detection Confusion Matrix. Each row of the confusion matrix represents a scene to target model library comparison. Each cell in a row shows the probability of detection that the target (with the ID shown in the top row) matches the scene (described at the beginning of the row). For each scene, the angular diversity and angular views is also shown in the first two columns to give a notional idea of the target coverage/obscuration.

From Table 5-1, we can see that the probability of detection confusion matrix resembles an identity matrix, which would be the ideal result. For all scene comparisons, the highest Pd value always falls on the target that matches the scene target truth. Furthermore, most of the remaining targets have a zero Pd because the recognition algorithm found no match between the respective target models and the scene. The rejection of a large portion of the candidate models in conjunction with most of the Pd falling on the correct target indicates that the recognition algorithm can readily discriminate the correct target from among the targets in the model library while achieving low false alarm rates.

In 8 out of the 12 scenes, the Pd fell almost entirely on the correct target at Pd levels exceeding 90%. For the remaining four data scenes, the correct target was still assigned the highest Pd, but a significant portion of the probability fell on targets other then the target truth. A closer examination of these four scenes reveals that while the Pd did not entirely fall on the correct target, the distribution of Pd values fell almost entirely on a single class of targets that included the target truth.

One such case is the BMP1 scene that matched the BMP1 model with a probability of 61% and the BMP2 model with a probability of 38%. Since the BMP1 and BMP2 targets have almost identical dimensions and spatial structure, the recognition algorithm was unable to discern the two models from one another. Nonetheless, the scene was recognized to contain a BMP-class vehicle with a probability of 98%. Thus, we can conclude that recognition algorithm was able to correctly classify the scene as a BMP with a probability of 98% and identify the target as a BMP-1 with a 61% probability.

Another example is the M1A1 scene, where Pd predominantly falls on two tanks: the M1A1 tank model at 83% and T72 tank model at 17%. A match of both tanks by the spin-image recognition algorithm is reasonable since tank-like targets are likely to have similar dimensions and spatial features, which in turn will result in a high correlation between the spin-images stacks of the targets. Even though the probability of detection did not all fall on the M1A1 model, we have 100% probability of detection that the scene is a tank. Thus, the recognition algorithm was able to classify the scene as a tank with a probability of 100% and identify the tank as an M1A1 with an 83% probability.

Another scene that demonstrates correct target classification is the Huntsville T72 scene, where the T72 tank model Pd is 87% while the M1A1 tank model Pd is 13%. Again, the recognition algorithm correctly classified the scene as a tank with 100% probability and identified the tank as a T72 with an 87% probability.

Overall, the probability of detection matrix shows that the recognition algorithm identified the correct target by assigning the largest Pd value for all twelve recognition tests. Considering Equation 5.2, which formally states that, each scene search is assigned the model with the maximum Pd value as the final identified target, we can conclude that we achieved a 100% recognition rate. Concurrently, the recognition algorithm rejected most of the other targets, which indicates good target discrimination and the potential for achieving very low false alarm rates. In addition, for all recognition tests, approximately 95% of the Pd fell on the correct class of targets, indicating a high level of discrimination between targets of different classes.

		Total # of Scene Points	% of scene points selected	Scene Resolution (meters)	Spin Image Size (#Pixels)	Stack Create Time (secs)	Avg. Match Time Per Model (secs)	Avg. Verify Time Per Model (secs)	Total Recognition Time Per Model (secs)
Field Data	**BMP-1** C5- F10-P3	12935	25%	0.13	100	15.6	178.74	16.9	197.2
	BTR-70 C5- F10-P4	13474	25%	0.127	100	19.2	197.2	5.48	204.6
	HMMV RMF May 02	6753	50%	0.14	100	9.40	146.19	0.62	147.75
	HMMV C8-F1-P10	5637	50%	0.13	100	7.83	123.59	0.31	124.68
	M1-A1 Eglin Dec 01	5603	50%	0.18	100	5.36	128.91	0.87	130.32
	M2-A3 C5-F13-P07	7209	50%	0.16	100	5.34	80.36	1.51	82.40
	M-35 C5-F10-P05	8753	50%	0.14	100	16.38	197.53	1.91	201.08
	M60 w/plow Huntsville May 02	3740	100%	0.17	100	5.82	95.30	0.16	96.04
	M60 Huntsville May 02	3676	100%	0.16	100	5.96	93.63	0.21	94.44
	M60 under camo C05-F16-P10	14402	25%	0.18	100	18.61	89.51	1.16	92.53
	T-72 C05 F0 P3	18093	25%	0.14	100	29.40	203.90	4.37	211.21
	T72 C20-F01-P3	13338	25%	0.13	100	30.60	115.16	3.47	121.69
	Average Total Time Per Model (seconds) =								**142.0**

Table 5-2: ATR Time Performance. Each row shows the recognition parameters and resulting timing statistics for a respective scene.

Table 5-2 above summarizes the recognition timing performance obtained for each of the twelve data sets shown above. The ATR algorithm was run on an Intel Pentium-4 Xeon running at 2GHz. In Table 5-2, the *Stack Create Time* column is the time taken to create the spin-image stack of the scene, the *Avg. Match Time* column is the average time used to match the scene spin-image stack to each model and generate pose transformations, and the *Avg. Verify Time* column is the average time taken to verify each scene to model comparison. The sum of the Stack Create Time, Avg. Match Time and Avg. Verify Time is shown in the *Total Recognition Time Per Model* column. Overall, we achieved a recognition time of approximately 2 minutes per model, which is close to our initial recognition goal of 1 minute per model.

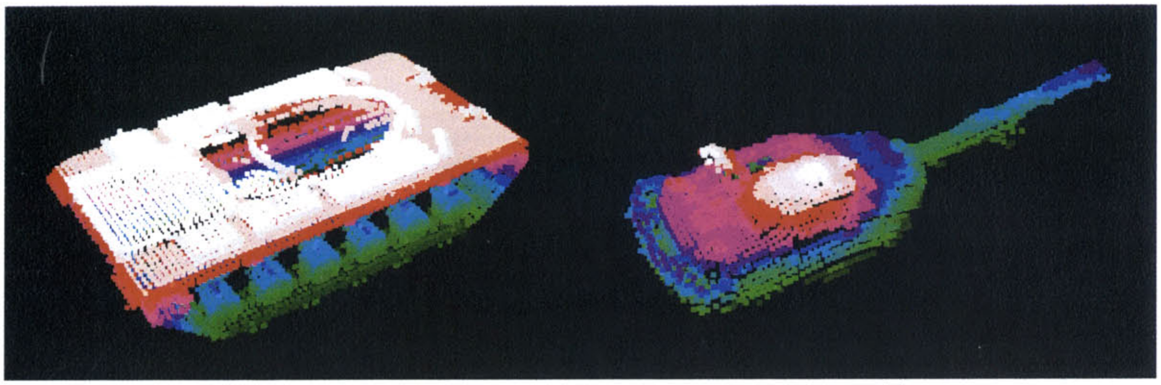
5.2.2 Articulated ATR Study

The recognition tests so far have dealt with targets that are represented by solid objects with no articulated components. We now want to extend the ATR algorithm to recognize articulated targets, with multiple movable parts that are in arbitrary orientations. The main benefit of articulated ATR is that we would have the ability to match an object regardless of the relative position of each of its movable parts (Ex: Tank with its turret rotated, Scud launcher with its missile at different angular pitches). Furthermore, recognition by parts allows the possibility of recognizing vehicles that come in many possible configurations, such as the multi-purpose HMMV platform and the myriad of one-of-a-kind technicals encountered in our current military campaigns. Another inherent benefit of articulated ATR is that we can also develop a higher level of tactical awareness by determining the current aim direction of a target's weapon.

5.2.2.1 Preliminary Results

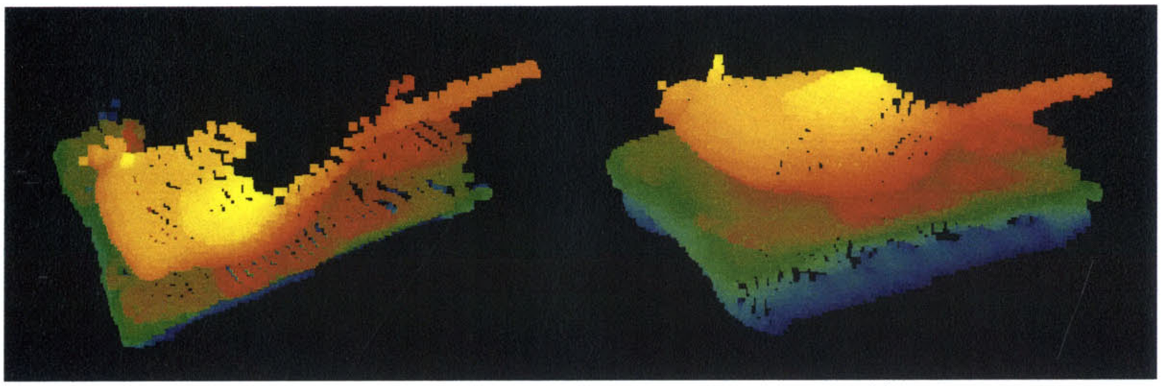
We ran a feasibility test to demonstrate articulated ATR on measured JIGSAW data. For the test, we created a model library containing M60 parts, namely an M60 tank body and an M60 tank turret. The spin-image library had a data resolution of 10cm, bin size of 12.5cm, support distance of 1.25 meters, and a support angle of 90-degrees. Figure 5-2 shows the parts model library.

The concept of articulated ATR was demonstrated on a scene containing a single-view measurement of an M60 tank with its turret turned by 180 degrees (see Figure 5-3). Figure 5-4 captures a qualitative summary of the results, showing that the correct pose transformation was found for each target part.



a) b)

Figure 5-2: Height color-coded M60 Tank parts. a) M60 body model. b) M60 turret model.



a) b)

Figure 5-3: Height color-coded single-view M60 tank with its turret rotated by 180 degrees. a) Orthographic view of scene. b) Sensor perspective view of the scene.

For the recognition of each part in the scene, the measured data present on the other target parts can be considered as clutter. For instance, in Figure 5-4-c and 5-4-d, when we are attempting to recognize the M60 turret in the scene, the measurements on the M60 body act as clutter. Even though the clutter from the M60 body is spatially adjacent to the M60 turret, the recognition algorithm is able to correctly identify the turret and compute a correct pose transformation. This successful recognition by parts shows the robustness of the spin-image algorithm to scene clutter, and its potential performance in the development of a fully articulated ATR system.

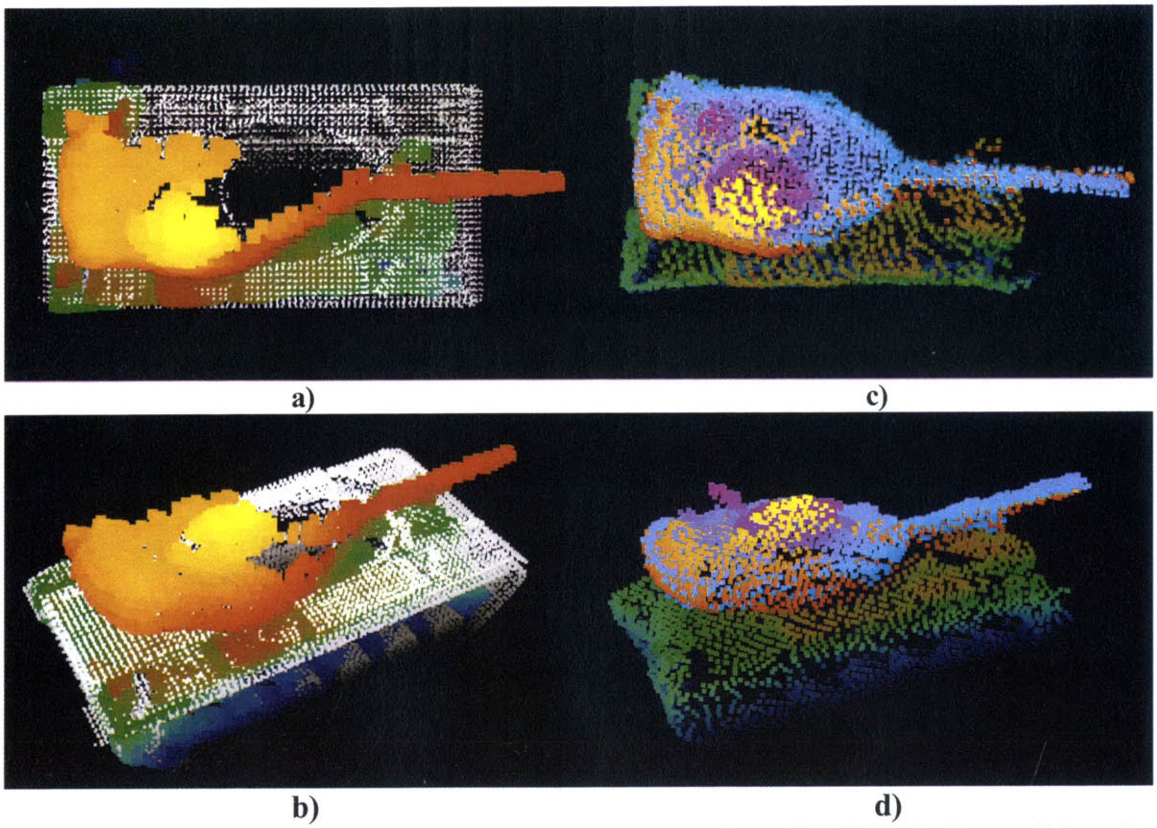


Figure 5-4: M60 Recognition by Parts. a) Orthographic view of M60 Body Recognition; the scene is height color coded in a green-red-yellow color map, while that M60 model body is height color coded as white points. b) Another perspective of the M60 body recognition to show that the correct pose was found in all six degrees of freedom. c) Orthographic view of the M60 turret recognition; the scene points are again height-color coded using a green-red-yellow color map, while the M60 turret model is height-color coded using a purple-blue color map. d) Another perspective of the M60 turret recognition to show that the correct pose was found in all six degrees of freedom.

In summary, we have thoroughly demonstrated good ATR system performance and shown the feasibility of pursuing articulated ATR. We have been able to achieve 100% recognition rate with Pd confidence measures that typically were above 90%. The high Pd levels indicates that the ATR algorithm can discriminate targets and has the potential for achieving very low false alarm rates. Furthermore, the distribution of the Pd values across the model library implies that the recognition algorithm can correctly classify targets based on similarities in the general target structure. In our results, approximately 96.4% of the total Pd measurement fell on the general target class that encompassed the target truth. In addition, the results of the study on Articulated ATR reiterated that spin-image matching is highly robust to occlusion and scene clutter in close proximity to the object of interest.

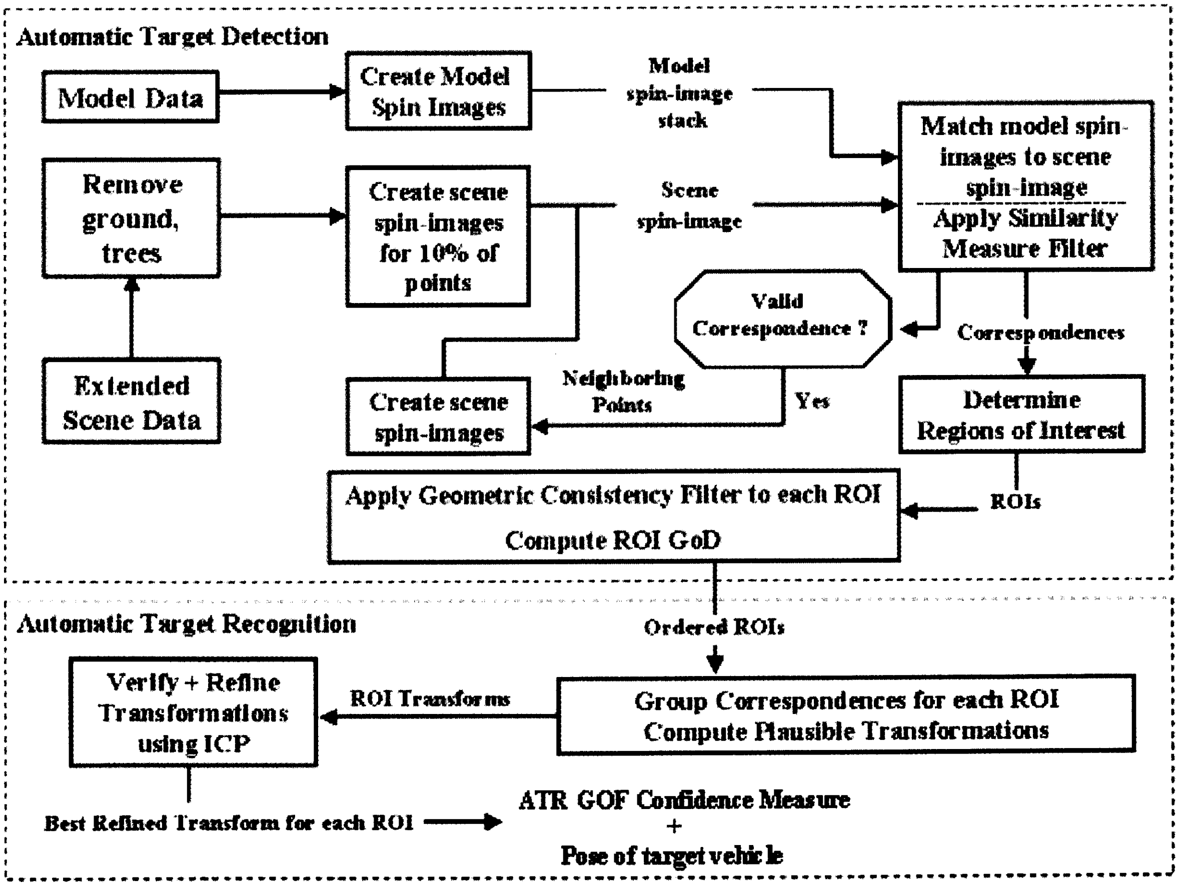
In the next chapter, we will combine our ATR algorithm with an automatic target detection algorithm and show the end-to-end performance of a fully automatic target detection and recognition system.

Chapter 6: Automatic Target Detection in Cluttered, Noisy Scenes

Automatic target detection (ATD) was performed using the general approach of "3D cueing," which determines and ranks regions of interest within a large-scale scene based on the likelihood that they contain the respective target. Spin-image matching is used to provide a statistical measure of the likelihood that a particular region within the scene contains the target [3]. The detection algorithm is based on the previous work of Hebert et al.

6.1 Detection Algorithm

The 3d-Cueing algorithm is tailored for target detection in large-scale terrain scenes. The implemented algorithm can detect and recognize multiple known targets in the scene.



Automatic Target Detection
Model Data
Create Model Spin Images
Model spin-image stack
Match model spin-images to scene spin-image
Apply Similarity Measure Filter
Remove ground, trees
Create scene spin-images for 10% of points
Scene spin-image
Valid Correspondence ?
Correspondences
Extended Scene Data
Neighboring Points
Yes
Create scene spin-images
Determine Regions of Interest
Apply Geometric Consistency Filter to each ROI
Compute ROI GoD
ROIs
Automatic Target Recognition
Ordered ROIs
Verify + Refine Transformations using ICP
ROI Transforms
Group Correspondences for each ROI
Compute Plausible Transformations
Best Refined Transform for each ROI
ATR GOF Confidence Measure
+
Pose of target vehicle

Figure 6-1: ATD-ATR Process Block Diagram

Figure 6-1 above shows a detailed block diagram of the ATD-ATR system for a scene to target model comparison. Similar to the spin-image matching algorithm described in Chapter 2, the detection algorithm starts with an input scene data set of oriented points. To reduce computation time, the search volume is constrained by removing the ground and canopy cover. Ground removal is based on the results of the ground registration algorithm described in Section 3.2.2. Given a known ground level from the registration results, the scene is cropped in height based on the maximum height of the respective target models. For our target model library, the height range is determined to be 0.5 to 4 meters above the detected ground.

A small fraction of points from the remaining oriented point data set is chosen to create corresponding spin-images. The scene data points are not judiciously picked: the points are uniformly distributed across the given scene. Therefore, no feature extraction was performed to pick certain points. Following Hebert's et al procedure, this fraction of points is set to 10% of the data set in order to reduce the computation time and allow the algorithm to be scalable to large-scale scenes. The fraction is typically large enough to have several points on the target of interest.

Corresponding spin-images are created for each chosen oriented point. The resulting spin-image stack is compared against a target model according to the spin-image matching correlation procedure described in Section 2.3.1. The correspondences found are then filtered using the similarity measure filter described in Section 2.3.2.

The remaining correspondences are used to create regions of interest within the scene. The process of creating regions of interest involves a recursive search for valid correspondences based on the location of the scene correspondences found so far. For each filtered correspondence, the closest neighboring points within the scene that have not already been checked are selected. Since only 10% of the points were checked so far, a large fraction of the points (i.e. 90%) remain untested by the spin-image correlation process. Spin-images are created for each of the closest oriented 3D points, correlated to the target model spin-image stack and filtered by the extreme outlier threshold. The remaining correspondences are filtered based on the similarity measure filter. For each new correspondence that passes the above filtering steps, its closest neighbors are analyzed. This recursive process stops when no more closest neighbors exist that pass the filtering procedure described above.

Given that a target exists within the scene, we expect several points on the target to be chosen in the initial fraction of 10%. When the spin-images of these particular points are compared to the spin-images of the target model, the correspondences formed should be able to pass the similarity measure filter. This will result in a recursion on the neighboring scene points, which are likely to be measurements of the target. The spin-image creation and correlation process will repeat itself, until all the closest neighbors that pass the filtering thresholds are found. Visually, this process will result in the growth of a group of correspondences that will define a target ROI. For our target detection experiments, the closest neighbor distance was set to 2 meters.

The ROIs obtained using the above algorithm can vary drastically in the number of correspondences, correspondence values and surface area coverage. To discriminate between the various ROIs, geometric consistency is used to remove unlikely correspondences. (see Section 2.3.2). Each ROI that passes the geometric consistency filter is rated with a goodness value that corresponds to its likelihood of matching the target of interest. The ROI goodness of detection value is defined as:

$$ROI\ GoD(s,m) = \frac{\mathrm{N}}{\mathrm{M}}\sum_{i=1}^{N} C_i \tag{6.1}$$

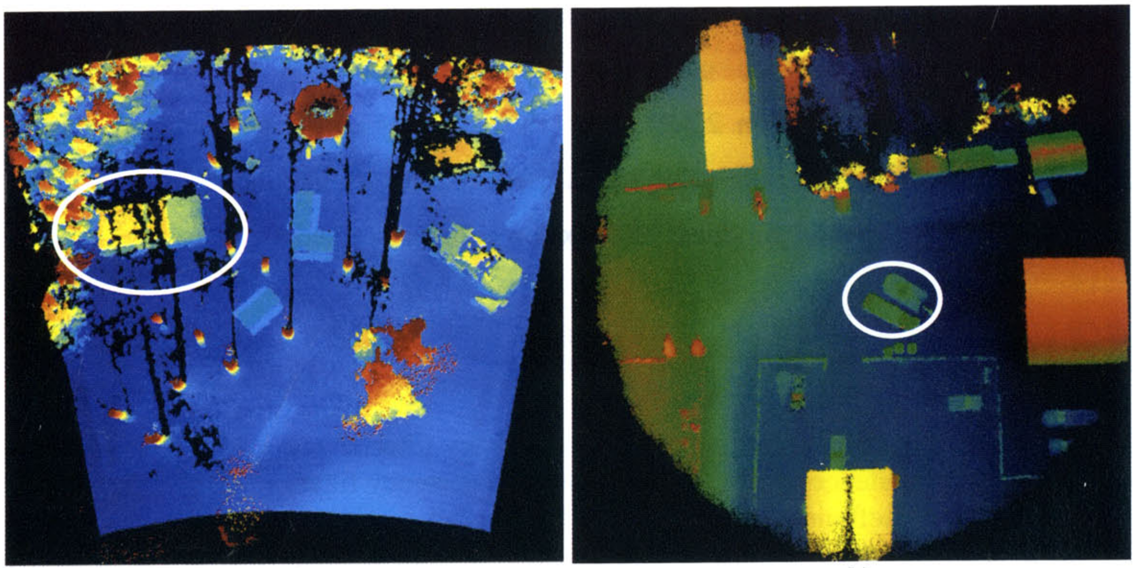
where
N = # of correspondences after geometric consistency filter
M = # of correspondences before the geometric consistency filter
Ci = Correlation coefficient value as defined by Eq 2.4

The ROIs are then queued based on their goodness of detection value. The ROI with the best GOD value is analyzed first by the recognition algorithm before proceeding to the second best ROI and so on. For each ROI, the correspondences are filtered and grouped according to Section 2.3.2. The resulting pose transformations are verified using the ICP algorithm and assigned a V_{GOF} value according to equation 2-11. The pose transformation with the best V_{GOF} is considered to be the final result of the scene ROI to model comparison.

The above ROI detection process is repeated for each target model. For each known target, a unique set of ROIs are found and analyzed to determine if a match exists. The end-result of the comparison of a scene to a library of models will be a list of ROIs, each matching a target model in a certain pose along with an ATR GoF that specifies the level of confidence that the match is correct. The ATR GoF confidence measure is equivalent to Pd as defined in equation 5.1.

6.2 Results & Discussion

Five extended terrain scenes recorded under the JIGSAW Phase-II data campaign were used to test the ATD-ATR system. Each data set contained one or more known targets and covered an area between 25x25 meters to 100x100 meters. Target truth in the form of GPS location and target ID was known prior to data collection. Targets in the data set were both out in the open and also underneath heavy canopy cover. Here is an orthographic view of the original data sets used for target detection:



a) **b)**

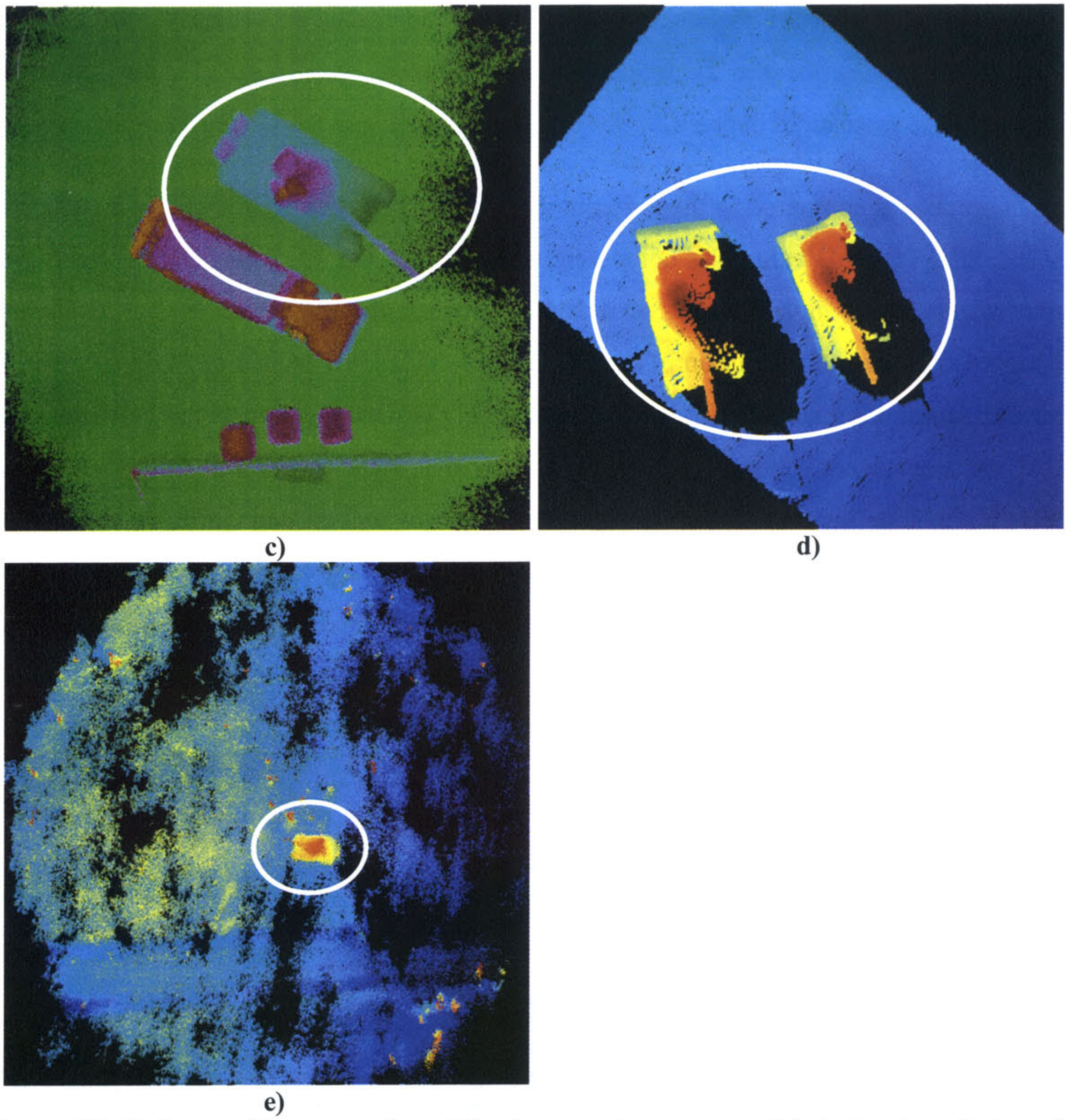


Figure 6-2: Orthographic perspective of five large-scale scenes used to test automatic target detection. For some of the data sets, the trees have been cropped out in order to show the obscured target. In each image, the white oval is used to pin point the location of the target of interest. a) 25x25 meter measured scene of an HMMW under canopy cover. b) 100x100 meter measured scene of T72 in a tank yard from a sensor altitude of 450 meters. c) 25x25 meter measured scene of a T72 in a tank yard from sensor altitude of 150 meters. d) 25x25 meter measured scene of two M60 tanks. e) 100x100 meter measured scene of a T72 underneath heavy canopy cover, from a sensor altitude of 450 meters.

Each scene was sub-sampled using 20cm voxels. The resolution down sampling was performed in order to reduce the computational complexity. To reduce computational complexity further, the spin-image resolution was also reduced from the 10x10 pixel spin-images used for recognition to spin-images with only 5x5 pixels. Support distance remained constant at 2.5 meters, while the bin-size increased accordingly from 25cm to 50 cm to account for the 2x reduction in spin-image resolution. The support angle remained the same at 90 degrees.

Each scene was compared to the target model library. For each ROI found in a particular scene, an ATG GoD value was computed using Eq 6.1. The ROI's ATD GoD value was normalized to the highest GoD value found between the scene and the target library. Figure 6-3 shows a distribution of the normalized ATD GoD values of ROIs found from all five tested scenes. The ROI distribution of ATD GoD values is divided between ROIs that were considered false alarms and the ones that were considered true positives. A false alarm is defined as a ROI that matches a target to background clutter or an ROI that incorrectly matches a known scene target to the wrong target model. A true positive is defined as an ROI found for a particular target model that encompasses the measurements of a scene target, whose target truth matches the respective target model.

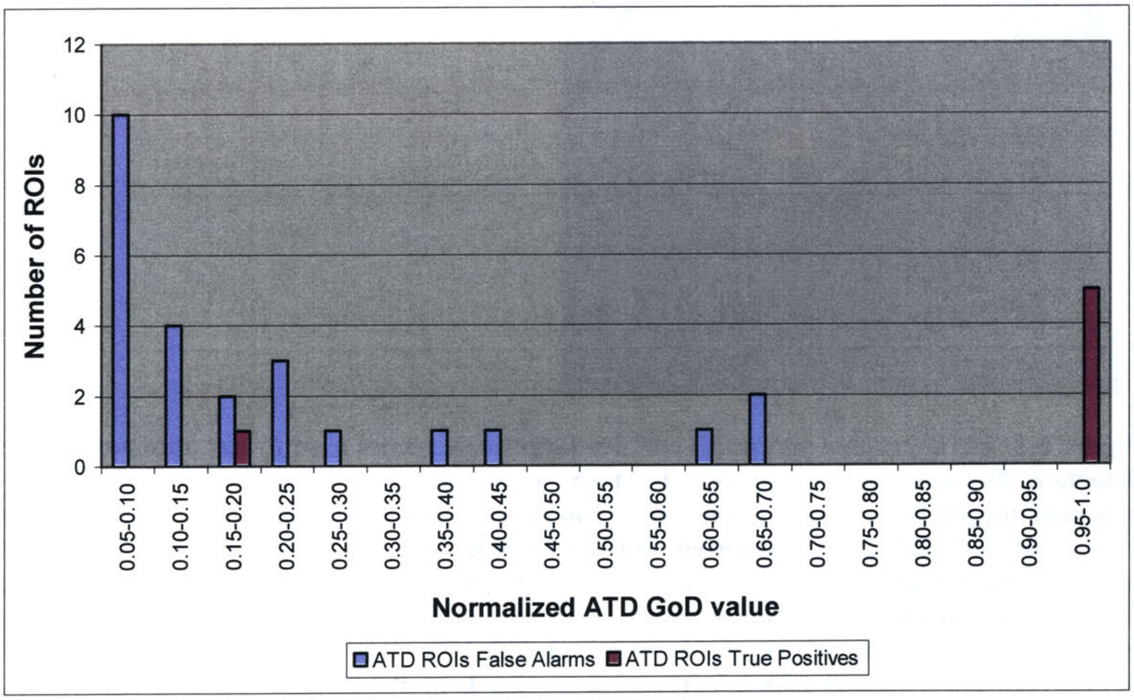


Figure 6-3: Normalized ATD GoD ROI Distribution

For all five scenes in Figure 6-3, a true positive ROI had the largest ATD GoD value, leading to a normalized ATD GoD value of 1. Thus for all five scenes, we were able to correctly detect and identify a target instance. The M60s scene presented an interesting case, where two identical M60-type targets existed within the scene. For this single-view scene, the ROI with the highest ATD GoD fell on the M60 target in the sensor's foreground; the second M60 was also detected, but with a much lower normalized ATD GoD value of 0.185. (corresponds in Figure 6-3 to the true positive under the 0.15-0.20 normalized ATD GoD bin) The large difference in GoD value for the two tanks in the scene is not surprising: the M60 tank in the sensor foreground had about 5318 measurements while the M60 tank further down in range from the sensor had about 3676 measurements. Since the surface-matching algorithm is dependent on the scene resolution for the creation, correlation and filtering of spin-images, a scene with variable resolution will result in a bias towards objects in the sensor's foreground, which are bound to have a higher spatial measurement resolution. Furthermore, the ATD GoD value is a function of the sum of point-correspondence values and is directly affected by the number of measurements on target. The M60s scene presents the following challenge in the detection of multiple instances of a target object within a scene: one of the detected object instances is bound to have a higher signal level than the rest, lowering the confidence that the rest of the objects are valid detections of the same target object. In our case, due to a lower resolution on the M60 that is down-range, the GoD confidence value is smaller than the GoD value of the foreground M60, thus lowering our confidence that the M60 tank in the background is a valid detection.

Ignoring the low-GoD true-positive result from the M60s scene, Figure 6-3 shows a good separation between the distributions of false alarms and true positives. The two distributions have a separation of approximately 0.33 in normalized ATD GoD space. This indicates that we can accurately detect and identify the correct target from background clutter and also identify the correct target from the library of known targets. With a separation of almost 1/3 of the GoD value space, a detection threshold can readily be set between the highest false alarm (at 0.671) and the lowest of the remaining true positives GoDs (at 1.0).

Thus, even as a stand-alone algorithm, the ATD works exceptionally well. We will now show the results of ATD coupled with ATR. Figure 6-4 show the distribution of normalized ATR GoF values obtained after we ran the ATR algorithm on the detected ROIs. From the distributions, we can discern that the most of the true positives mapped to a normalized ATR GoF value of 1. Again, the multiple M60s targets presented a challenge with the background M60 tank mapping to a normalized ATR GoF of 0.24, slightly higher than the 0.18 ATD GoD value. There is also a significant improvement in the distribution of false alarms and true positive in the ATR GoF space as compared to the ATD GoD space. Most of the ATD false alarms have been remapped from an ATD GoD range of 0 to 0.67 to an ATR GoF range of 0 to 0.24. The re-mapping of false alarms to lower ATR GoF values further increases the separation between the distribution of false alarms and true positives. The larger separation between false alarms and true positives represents an improvement in our ability to discern the correct target from background clutter and other known targets. Therefore, the ATR value space is an improvement over the ATD value space.

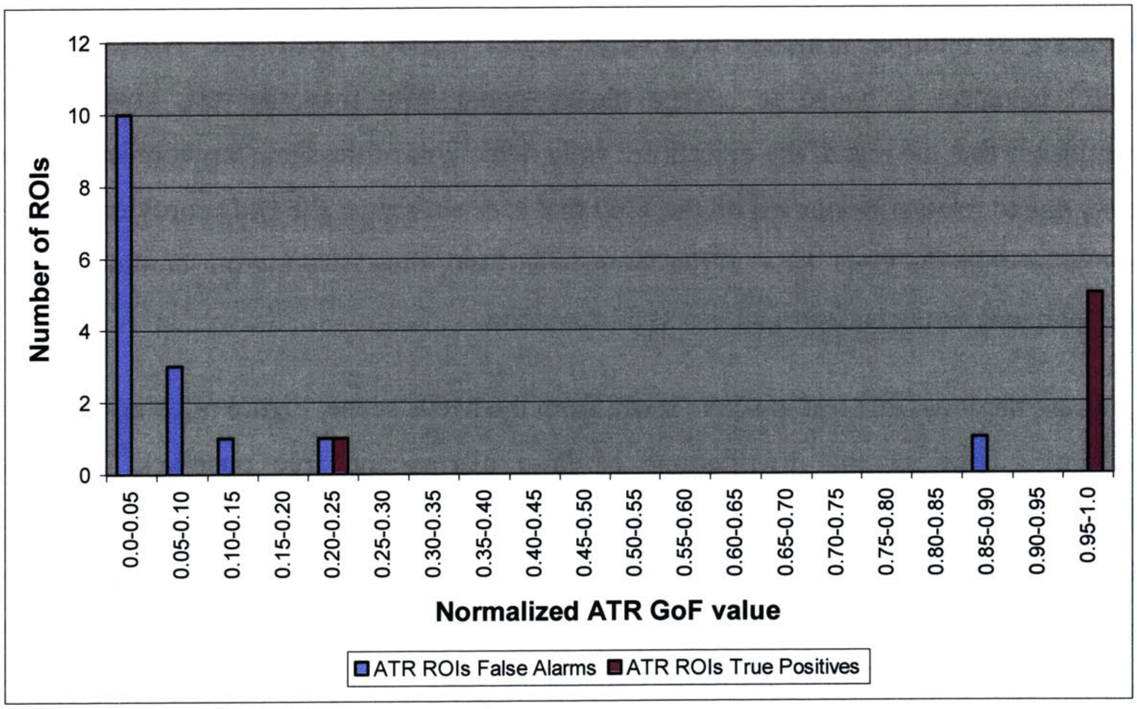


Figure 6-4: Normalized ATR GoF ROI Distribution

Table 6-1 shows the time performance of the entire ATD and ATR system. The ATD-ATR system was run on an Intel Pentium-4 Xeon at 2GHz. In Table 6-1, *Stack Create Time* is the time taken to create the spin-image stack of the scene. *The Average ATD+ATR Time Per Model* is the time used to detect ROIs for a model, and recognize whether the ROI is a valid target model instance. *The Average ATD+ATR Time Per*

Model also includes the contribution of the time taken to create the scene spin-image stack, weighted down by the number of models in the library, since the scene stack is computed only once and used for all the following target model comparisons. The last column in Table 6-1 is the *Average Detection + Recognition Time per Model* as a function of real time, where real time is defined as the data collection time. Overall, we achieved a recognition time of approximately 1.5 minutes per model, which translates into 9X real time.

	Scene Description	Total # of Scene Points (before ground removal)	Total # of Scene Points (after ground removal)	Target % of scene by volume (based on original scene with ground + canopy)	% of scene points selected to correlate to models	Scene Resolution (meters)	Spin Image Size (#Pixels)	Spin image Stack Create (seconds)	Avg ATD + ATR Time per model (based on the 10 model library) (seonds)	Avg ATD+ATR time per model versus real time
Field Data	**HMMV Scene** Huntsville June-03 C8-F1-P10	192097	26318	0.76	100%	0.25	25 (5x5)	59.8	120.72	N.A.
	M60s Scene Eglin 01	48997	8995	9.18	100%	0.16	100 (10x10)	15.6	497.86	N.A.
	Tank yard (450m alt) Huntsville Dec 02 C20-F02-P05	575938	35157	0.59	100%	0.26	25 (5x5)	40.9	219.26	18.66
	T72 under canopy (450m altitude) Huntsville Dec 02 C20-F02-P07	312189	10293	2.41	100%	0.18	25 (5x5)	29.9	45.33	5.04
	Tank yard (150m alt) Huntsville, Dec 02 C20-F01-P3	32750	7286	8.99	100%	0.24	25 (5x5)	7.62	30.69	3.32
Avg. ATD+ATR Time for 20 cm sub-sampled scenes (in seconds) =									**104.00**	
Average ATD+ATR Time per Model versus Real-time =										**9.01 X**

Table 6-1: ATD & ATR System Time Performance

In summary, our new ATD+ATR algorithm has demonstrated close to real time performance and good detection and identification accuracy. Given its timing and accuracy performance, the ATD+ATR system may have significant practical value to a human operator for aided target recognition under battlefield conditions.

Chapter 7: Conclusion

In this thesis research, we developed and implemented a fully automated target detection and recognition system that uses geometric shape and size signatures from target models to detect and recognize targets under heavy canopy and camouflage cover in extended terrain scenes. In support of this ATD-ATR system, we have also developed a novel method for data integration to register multiple scene views and obtain a more complete 3D surface signature of a target.

The ATD-ATR system performance was demonstrated on five measured scenes with targets both out in the open and under heavy canopy cover, where the target occupied between 1 to 10% of the scene by volume. The ATR section of the system was successfully demonstrated for twelve measured data scenes with targets both out in the open and under heavy canopy and camouflage cover. Correct target identification was also demonstrated for targets with multiple movable parts that are in arbitrary orientations. We achieved a high recognition rate (over 99%) along with a low false alarm rate (less than 0.01%).

The major contribution of this thesis is that we proved that spin-image-based detection and recognition is feasible for terrain data collected in the field with a sensor that may be used in a tactical situation. We also demonstrated recognition of articulated objects, with multiple movable parts. Considering the timing and accuracy performance, the ATD-ATR system may have significant practical value to a human operator for aided target recognition under battlefield conditions.

Immediate benefits of the presented work will be to the area of Automatic Target Recognition of military ground vehicles, where the vehicles of interest may include articulated components with variable position relative to the body, and come in many possible configurations. Other application areas include human detection and recognition for Homeland Security, and registration of large or extended terrain scenes.

Bibliography

[1] Ratches, J.A.; Walters, C.P.; Buser, R.G.; Guenther, B.D., **Aided and automatic target recognition based upon sensory inputs from image forming systems**, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 19 Issue: 9, Page(s): 1004 –1019, 1997

[2] Wellfare, M.; Norris-Zachery, K. **Characterization of articulated vehicles using ladar seekers,** Proceedings of the SPIE - The International Society for Optical Engineering, vol.3065 p. 244-54

[3] *O. Carmichael and M. Hebert,* **3D Cueing: A Data Filter For Object Recognition**, *IEEE Conference on Robotics and Automation (ICRA '99)*, Vol. 2, May, 1999, pp. 944 - 950.

[4] Richard M. Heinrichs, Brian F. Aull, Richard M. Marino, Daniel G. Fouche, Alexander K. McIntosh, John J. Zayhowski, Timothy Stephens, Michael E. O'Brien, Marius A. Albota, **Three-Dimensional Laser Radar with APD Arrays** Proceedings of the SPIE - The International Society for Optical Engineering Vol. #4377-14

[5] Zheng, Q.; Der, S., **Model-based target recognition in ladar imagery** Proceedings of the SPIE - The International Society for Optical Engineering, vol.3380 p. 343-51

[6] Dufour, J.-Y.; Martin, V. **Active/passive cooperative image segmentation for automatic target recognition,** Proceedings of the SPIE - The International Society for Optical Engineering vol.2298 p. 552-60

[7] Arnold, G.D.; Sturtz, K.; Weiss, I., **Detection and recognition in LADAR using invariants and covariants,** Proceedings of the SPIE - The International Society for Optical Engineering vol.4379 p. 25-34, 2001

[8] Zhou, Yi-Tong, Sapounas, Demetrios, **An IR/LADAR automatic object recognition system,** Proceedings of the SPIE - The International Society for Optical Engineering Vol. 3069, p. 119-128, 1997

[9] S. Grossberg and L. Wyse, **A Neural Network architecture for figure-ground separation of connected scenic figures**, Neural Networks, pp 723-742, 1991

[10] Zhengrong Ying and David Castanon, **Statistical Model For Occluded Object Recognition,** Proceedings of the 1999 International Conference on Information Intelligence and Systems, Page(s): 324 -327

[11] Sadjadi, F., **Application of genetic algorithm for automatic recognition of partially occluded objects,** Proceedings of the SPIE - The International Society for Optical Engineering vol.2234 p. 428-34, 1994

[12] Khabou, Mohamed A.; Gader, Paul D.; Keller, James M., **LADAR target detection using morphological shared-weight neural networks,** Machine Vision and Applications v 11 n 6 Apr 2000. p 300-305, 2000

[13] M. Hebert and J. Ponce, **A new Method for segmenting 3-D Scenes into Primitives**, SPIE Vol 1222 Laser Radar V, pp. 2-23, 1990

[14] R. Hoffman and A. Jain, **Segmentation and Classification of range images**, IEEE Transactions. PAMI, Vol. 9, No. 5 pp. 608-620, 1987

[15] D. Milgram and C. Bjorklund, **Range Image Processing: planar surface extraction**, SPIE Vol 783 Laser Radar II, pp 109-122, 1987

[16] Jianbing Huang and Chia-Hsiang Menq, **Automatic Data Segmentation for Geometric Feature Extraction from Unorganized 3-D Coordinate Points**, IEEE Transactions on Robotics and Automation, Vol 17, No. 3, 2001

[17] In Kyu Park; Il Dong Yun; Sang Uk Lee, **Automatic 3-D model synthesis from measured range data**, IEEE Transactions on, Volume: 10 Issue: 2, Page(s): 293 –301, 2000

[18] Cobzas, D.; Zhang, H., **Planar patch extraction with noisy depth data**, Third International Conference on 3-D Digital Imaging and Modeling, page(s): 240 –245, 2001

[19] Stein, F.; Medioni, G, **Structural indexing: efficient 3D object recognition**, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 14, Issue: 2, Page(s): 125 -145, 1992

[20] O. Carmichael, D.F. Huber, and M. Hebert, **Large Data Sets and Confusing Scenes in 3-D surface Matching and Recognition**, *Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling (3DIM'99)*, October, 1999, pp. 358-367.

[21] Shantaram, V.; Hanmandlu, M., **Contour based matching technique for 3D object recognition**, Proceedings. International Conference on Information Technology: Coding and Computing, page(s): 274 –279, 2002

[22] Dorai, C.; Jain, A.K., **Shape spectra based view grouping for free-form objects**, Proceedings, International Conference on Image Processing, Volume: 3, Page(s): 340 -343 vol.3, 1995

[23] Yamany, S.; Farag, A., **3D objects coding and recognition using surface signatures**, Proceedings of the 15th International Conference on Pattern Recognition, Volume: 4, page(s): 571 –574, 2000

[24] A. Johnson, **Spin-Images: A Representation for 3-D Surface Matching**, Ph.D. Thesis in Robotics at the Robotics Institute, Carnegie Mellon University

[25] A. Johnson and M. Hebert, **Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes**, Proceeding of IEEE Conference on Pattern Analysis and Machine Intelligence, Vol. 21, No. 5, May 1999

[26] Andrew E. Johnson and Martial Hebert., **Surface matching for object recognition in complex three-dimensional scenes.**, *Image and Vision Computing,* **16**, pp. 635-651, 1998

[27] A. Johnson, O. Carmichael, D.F. Huber, and M. Hebert, **Toward a General 3-D Matching Engine: Multiple Models, Complex Scenes, and Efficient Data Filtering**, Proceedings of the 1998 Image Understanding Workshop (IUW '98), November, 1998, pp. 1097-1107.

[28] Hodge, J.; DeKruger, D.; Park, A. **Mobile target LADAR ATR system,** Proc. SPIE - Int. Soc. Opt. Eng. (USA) vol.4379 p. 35-50, 2001

[29] Cook, T.D. **Results of Ladar ATR captive flight testing experiments,** Proc. SPIE - Int. Soc. Opt. Eng. (USA) vol.4379 p. 78-85, 2001

[30] C. Chua and R. Jarvis. **3-D free-form surface registration and object recognition**. *Int'l Jour. Computer Vision*, vol. 17, no. 1, pp. 77-99, 1996.

[31] B. Horn. **Closed-form solution of absolute orientation using unit quaternions**. *Jour. Optical Society of America*, vol. 4, no. 4, pp. 629-642, 1987.

[32] P. Besl and N. McKay. **A method of registration of 3-D shapes**. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 2, pp. 239-256, February 1992.

[33] Y. Chen and G. Medioni. **Object modelling by registration of multiple range images.** *Image and Vision Computing*, vol. 10, no. 3, pp. 145-155, 1992.

[34] Z. Zhang. **Iterative point matching for registration of free-form curves and surfaces**. *Int'l Jour. Computer Vision*, vol. 13, no. 2, pp. 119-152, 1994.

[35] R. M. Marino, W. R. Davis, Jr., G. C. Rich, J. L. McLaughlin, A. G. Vogel, B. B. Chen, D. E. Weidler, G. Rowe, R. E. Hatch, L. J. Skelly, T. E. Square, M. E. O'Brien, J. W. Burnside, B. M. Stanley, E. I. Lee, D. J. Ruscak, B. F. Aull, J. J. Zayhowski, **Jigsaw 3D imaging ladar with a photon-counting Geiger-mode APD array: sensor and measurements,** Proceedings of SPIE, Laser Radar Technology and Applications VIII, vol. 5086, pp. xx-xx, 2003.

[36] Schroeder, J.; Extended Abstract on **Automatic Target Detection and Recognition Using Synthetic Aperture Radar Imagery**, Cooperative Research Centre for Sensor Signal and Information processing (CSSIP) SPRI building, Mawson Lakes Boulevard Mawson Lakes, SA 5095