Object to Multisensor Coregistration with Eight Degrees of Freedom[*][†]

Anthony N. A. Schwickerath and **J. Ross Beveridge**
Computer Science Department
Colorado State University
schwicke/ross@cs.colostate.edu

Abstract

A new least-squares procedure is presented
which fuses data from a standard perspective
sensor (CCD camera, FLIR sensor, etc.) and a
range sensor (LADAR) based upon correspond-
ing features identified on a 3D object model
and in each image. The algorithm solves for
both the pose estimate of the object relative to
the sensors and the registration between sen-
sors. This model-based coregistration process
is being developed to support future work rec-
ognizing modeled 3D objects in scenes imaged
by both optical (FLIR and CCD) and range
(LADAR) sensors. Coregistration results are
presented for both synthetic and real world
tests. The algorithm requires an initial pose
and sensor registration estimate. Tests on con-
trolled synthetic data show it is robust with re-
spect to substantial errors in initial translation
and orientation errors up to roughly 45 degrees.

1 Introduction

The work presented here is part of a larger project de-
veloping new ways of identifying modeled 3D objects in
range, IR and color data [Beveridge *et al.*, 1994a]. The
project task domain is the recognition of military ve-
hicles from an Unmanned Ground Vehicle performing
a Reconnaissance, Surveillance and Target Acquisition
(RSTA) task. The project goal is the development of
recognition algorithms which extend the current state-
of-the-art for occluded targets, targets viewed against
structured or cluttered backgrounds, and targets viewed
at uncharacteristic angles.

The approach being taken by this project is to de-
velop new algorithms which precisely match 3D object
models to range and optical data. The approach to
matching will be similar to those previously developed
for single CCD camera problems [Beveridge *et al.*, 1990;
Beveridge, 1993; Beveridge and Riseman, 1994].

The term *coregistration* signifies that matching will si-
multaneously refine both the estimated 3D pose of the
object relative to the sensor as well as the registration
parameters relating the coordinate systems of the two
sensors. The first step toward coregistration matching
is the development of coregistration algorithms capable
of determining object pose and sensor-to-sensor registra-
tion.

In setting up the coregistration problem, we assume
that intrinsic parameters for range, color and IR sensors
are known. Moreover, we assume that corresponding
point and line features are already identified on a 3D
object model and in both a range and an optical image.
The range image is assumed to directly measure distance
to points in the scene. The projection model for the op-
tical senor is perspective projection. The optical sensor
may be either a color CCD camera or a FLIR sensor.

An eight degree of freedom coregistration problem is
addressed in this paper, where six degrees of freedom re-
late the object to the optical camera, and two relate the
range sensor to the optical sensor. The six degrees of
freedom allow the object free movement relative to the
two sensors. The two degrees of freedom allow trans-
lation between sensors in a common image plane. This
translation compensates for small errors in the initial
registration between range and optical imagery.

The goal of coregistration is the best estimate of
the eight coregistration parameters. This best esti-
mate minimizes a squared distance function between
corresponding object and image features. The prob-
lem setup is an extension of a single optical sensor
pose algorithm developed by Kumar [Kumar, 1989;
Kumar and Hanson, 1994]. Kumar's algorithm itera-
tively solves the non-linear optimization problem associ-
ated with minimizing a sum of point-to-plane distances.
Typically the points are endpoints of 3D line segments
on the object model. The planes are typically defined
by the optical sensor focal point and endpoints of corre-
sponding line segments in the image plane.

In this work, Kumar's error function is generalized to
include a sum of squared Euclidean distances between
corresponding range and model points. Additionally,
two more degrees of freedom are added to the problem
to permit translation of the range sensor relative to the
optical sensor. These distance measures are shown in
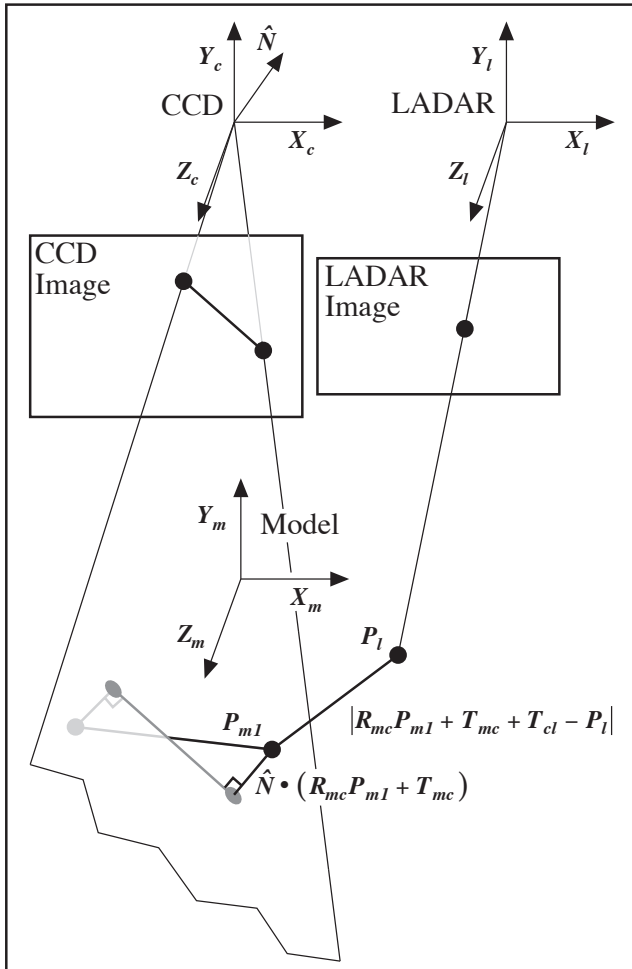Figure 1 and are more thoroughly explained in Section 4.

Figure 1: Illustrating distance errors which define optimal coregistration.

2 Background

A long tradition of work on object recognition has emphasized finding matches between object and image features such that there is a single globally consistent alignment of features. For example, Lowe utilized this approach subject to orthographic projection [Lowe and Binford, 1985] and later demonstrated object tracking under perspective projection [Lowe, 1991]. Huttenlocher [Huttenlocher and Ullman, 1990] demonstrated alignment based recognition under orthographic projection. Grimson's work [Grimson, 1990] on constraint base matching has emphasized local feature compatibility to prune tree search and thus minimize tests of global alignment.

Our own previous work [Beveridge *et al.*, 1990; Beveridge *et al.*, 1991; Beveridge, 1993] emphasizes global alignment as a basis for match ranking and optimal matching. Using work of Kumar [Kumar, 1989; Kumar and Hanson, 1994], this approach generalizes to matching subject to perspective given approximate estimates of initial object pose [Beveridge and Riseman, 1992; Beveridge and Riseman, 1994]. The coregistration work presented here is the next logical step in a generalization

of our optimal matching work to multisensor problems. Kumar's single optical sensor pose algorithm is extended to include additional data from a range sensor and additional degrees of freedom associated with registration between the sensors.

Coregistration is a form of model-based sensor fusion, and as such relates closely to previous work on multisensor fusion. Aggarwal [Aggarwal, 1990] summarizes past sensor fusion work and makes two points particularly relevant to this paper. Aggarwal notes that past work on sensor fusion emphasized single modality sensors, with comparatively little work on different sensor modalities. The implied explanation is that relating data from different modalities is more difficult. While Aggarwal [Magee *et al.*, 1985] and others [Stentz and Goto, 1987] have examples of successful mixed modality fusion, this is still a young research area.

Aggarwal also notes that to properly perform mixed modality sensor fusion, co-ordinate transformations between images need to be adaptively determined. The scene dependence of these transformations sets the stage for the model-based coregistration work developed in this paper. We believe this dependence upon scene geometry implies that known sensor and object geometry should be used to constrain and adapt both the object pose and sensor-to-sensor co-ordinate transformations as part of object recognition.

Others have worked on problems similar to the coregistration problem discussed in this paper. Herbert [Hebert *et al.*, 1990] presents a clean least-squares mechanism for computing the rigid transform between a range and color CCD sensor based upon corresponding image points in the two sensor images. However, this work does not recover explicitly the associated pose of the 3D points relative to the sensors. Both Eason [Eason and Gonzalez, 1992] and Hel-Or [Hel-Or and Werman, 1993] develop least-squares multisensor pose algorithms. However, these algorithms only solve for the six degree of freedom pose estimate. They do not support simultaneous adjustment of the sensor registration parameters. More recent work by Hel-Or and Werman [Y. Hel-Or and M. Werman, 1994] adds degrees of freedom to account for articulated objects and nicely handles variations of constraints in a single extended Kalman filter formulation.

3 RSTA Motivation

It is fair to ask why an automated RSTA system need compute the precise coregistration between object models and sensors. Typically, the end goal of RSTA is reliable recognition, and there are simpler ways of fusing multisensor data which do not include coregistration. For example, one sensor might be used for queuing and a second for confirmation. Alternatively, information may be independently extracted from multiple images and combined using Bayesian statistics or Schafer Dempster evidence accumulation. The point is, what information follows from coregistration which simply cannot be obtained using these other approaches?

The answer, in general terms, is that any constraint derived from precise knowledge of relative geometry requires coregistration. Perhaps the most obvious such

constraint is that the object model globally fit corresponding sensor features. This fit is naturally expressed as a function of the residual squared-error after coregistration. Another constraint is that sensor features not be missing or omitted from the match. Omitted features are best detected with the object model placed in the best geometric registration relative to the sensor data. Finally, coregistration allows constraints which involve sophisticated dependencies.

To illustrate this last point, consider the problem of occlusion and how detection in LADAR can permit proper handling in coregistered FLIR or CCD data. Surface occlusion is potentially determined using stereo [Bolles et al., 1993] or active vision [Kutulakos and Dyer, 1994]. However, it is essentially impossible to detect from a single IR or color image. Thus, using single IR or color images in a RSTA scenario it is impossible to define a match evaluation measure which can distinguish between occlusion and omission. Occlusion here refers to the case in which an occluding surface explains missing features. This contrasts with omission, in which features expected to be visible are inexplicably missing.

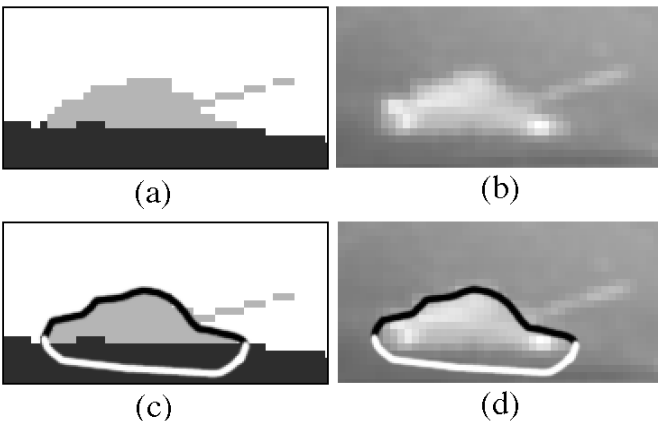(a)   (b)

(c)   (d)

Figure 2: Occlusion evident in LADAR can modify FLIR expectations. a) LADAR, b) FLIR, c) contour coded to indicate occlusion, d) contour showed in relation to FLIR.

Figure 2 illustrates how occlusion might be detected in LADAR and the result applied to determining what portion of a bounding contour would be expected to be visible in an IR image. Figure 2a shows a portion of the Fort Carson LADAR image [1] nov31205l gated to show the foreground in dark grey, the tank in light grey and the background in white. A corresponding portion of the FLIR image nov31553f is shown in Figure 2b.

Figure 2c shows a contour of the entire tank with the non-occluded portion shown in black and the occluded portion shown in white. Note that determining the occluded portion is a simple matter of testing what portion of the contour lies over the dark grey foreground area. The same contour is shown over the FLIR data in Figure 2d. Given precise coregistration, the lack of a lower

[1]See [Beveridge et al., 1994b] for a description of the Fort Carson data.

contour for the tank in the FLIR is explained by the occlusion evident in the LADAR. A several pixel error in registration between the FLIR and LADAR would make this type of reasoning between images and between images and object models impossible.

4 Coregistration Derivation

The coregistration procedure derived here assumes known 3D object features, a perspective sensor modeled as a pin-hole camera, and a range sensor. Although the derivation holds for any pair of perspective and range sensors, for simplicity, the perspective sensor is designated as the CCD and the range sensor as the LADAR.

Corresponding point or line features on the object and in the CCD image must be provided along with corresponding point features on the object model and in the LADAR image. The optimal set of coregistration parameters jointly minimize a sum-of-squared distances, an error, defined with respect to these corresponding features.

$$E_{tot} = w_{mc}E_{mc} + w_{ml}E_{ml} \qquad (1)$$

The first term, E_{mc}, measures distance between corresponding CCD and object features. This term is precisely the point-to-plane error criterion defined by Kumar [Kumar, 1989; Kumar and Hanson, 1994] for computing camera to object pose. The second, E_{mcl}, is simply the sum-of-squared Euclidean distances between corresponding object and range points. The subscripting convention orders letters indicating coordinate systems from left to right, with m standing for model, c standing for CCD, and l standing for LADAR. For instance, E_{ml} is the error for the mapping model-to-LADAR (ml).

The model-to-CCD error, E_{mc}, measures the distance between endpoints of 3D line segments on the object model and 3D planes defined by corresponding line segments found in the CCD image. These planes are defined by three points: the two endpoints of the line segment and the focal point of the CCD camera. The model-to-CCD error may be written as:

$$E_{mc} = \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\hat{N}_{ci} \cdot \left(R_{mc}\vec{P}_{mij} + \vec{T}_{mc} \right) \right)^2 \qquad (2)$$

where $\hat{N}_{ci}$ is the normal to the plane defined by the ith image line, R_{mc} is the rotation from model to CCD coordinates, $\vec{P}_{mij}$ is the jth endpoint of the ith model line segment in model coordinates, and $\vec{T}_{mc}$ is the translation from model to CCD coordinates. The weighting term λ_{ci} is typically 1, but can be used to bias some features over others.

For LADAR, 3D Cartesian points are readily determined for each pixel. Each Cartesian point is obtained by back-projecting into the scene by the measured range value along a 3D ray defined by the LADAR focal point and the pixel position on the LADAR image plane. The model-to-LADAR error, E_{ml}, is defined to be the squared Euclidean distance between these points and corresponding model points. It may be written as:

$$E_{ml} = \sum_{i=1}^{n_l} \lambda_{li} \left| \left(\left(\vec{P}_{mci} + \vec{T}_{cl} \right) - \vec{P}_{li} \right) \right|^2 \qquad (3)$$

$$\vec{P}_{mci} \;\;=\;\; R_{mc}\vec{P}_{mi} + \vec{T}_{mc}$$

where $\vec{P}_{mci}$ is the ith model point mapped into CCD coordinates, $\vec{T}_{cl}$ is a translation from CCD to LADAR coordinates which allows one image plane to translate relative to the other, and $\vec{P}_{li}$ is the corresponding ith measured LADAR point. The constituent parts of the total coregistration error are illustrated in Figure 1.

Minimizing equation 1 relative to the 9 terms in R_{mc} independently would violate the condition that R_{mc} be is a rotation matrix. To overcome this problem, Rodriguez's formula approximates small rotations with a cross product and is used here to linearize the rotation about the current estimated rotation R_{mc}^e. The change in rotation relative to the current estimate is represented by the vector $\delta\vec{\omega}_{mc}$.

$$\begin{aligned} R_{mc}\vec{P}_{mi} &= R_{mc}^e\vec{P}_{mi} + \delta\vec{\omega}_{mc} \times \left(R_{mc}^e\vec{P}_{mi} \right) \\ &= \vec{P}_{mi}^e + \delta\vec{\omega} \times \vec{P}_{mi}^e \end{aligned} \qquad (4)$$

To provide a more compact notation the vector $\vec{P}_{mi}^e$ is introduced and the subscript is dropped from $\delta\vec{\omega}$.

The error terms in equations 2 and 3 may now be rewritten as follows.

$$E_{mc} = \sum_{i=1}^{n_c}\sum_{j=1}^{2} \lambda_{ci}\left(\hat{N}_{ci} \cdot \left(\vec{P}_{mij}^e + \delta\vec{\omega} \times \vec{P}_{mij}^e + \vec{V} \right) \right)^2$$

where

$$\vec{V} = \vec{T}_{mc}^e + \Delta\vec{T}_{mc}$$

$$(5)$$

$$E_{mcl} = \sum_{i=1}^{n_l} \lambda_{li}\left| \left(\vec{P}_{mi}^e + \delta\vec{\omega} \times \vec{P}_{mi}^e + \vec{V} - \vec{P}_{li} \right) \right|^2$$

where

$$\vec{V} = \vec{T}_{mc}^e + \Delta\vec{T}_{mc} + \vec{T}_{cl}^e + \Delta\vec{T}_{cl}$$

In order to minimize E_{tot} in equation 1 with respect to $\delta\vec{\omega}$, $\Delta\vec{T}_{mc}$ and $\Delta\vec{T}_{cl}$, the partial derivatives with respect to each are set to zero and the resulting system of equations solved for coregistration update parameters. These partial derivatives are shown in Table 1. The following matrix M_{mci} is introduced to simplify the expressions.

$$M_{mci} = \begin{pmatrix} 0 & \left(\vec{P}_{mi}^e\right)_z & -\left(\vec{P}_{mi}^e\right)_y \\ -\left(\vec{P}_{mi}^e\right)_z & 0 & \left(\vec{P}_{mi}^e\right)_x \\ \left(\vec{P}_{mi}^e\right)_y & -\left(\vec{P}_{mi}^e\right)_x & 0 \end{pmatrix} \qquad (6)$$

The matrix M_{mci} is the righthanded vector product with $\vec{P}_{mi}^e$. It is also the partial derivative of $\delta\vec{\omega} \times \vec{P}_{mi}^e$ with respect to $\delta\vec{\omega}$.

The three equations in Table 1 constitute 9 linear equations in 9 unknowns. These linear equations may be written as:

$$\begin{pmatrix} A & B & C \\ D & E & F \\ G & H & J \end{pmatrix} \begin{pmatrix} \delta\vec{\omega}_{mc} \\ \Delta\vec{T}_{mc} \\ \Delta\vec{T}_{cl} \end{pmatrix} = \begin{pmatrix} \vec{K} \\ \vec{L} \\ \vec{M} \end{pmatrix} \qquad (10)$$

where the constants A through M are defined in Table 2.

For the coregistration case desired here, LADAR is not to be allowed to translate ahead of or behind the CCD sensor. Therefore, the CCD to LADAR translation is set to $\begin{pmatrix} t_x & t_y & 0 \end{pmatrix}^T$ and the rightmost column and bottom row of the 9×9 matrix in equation 6 are dropped. The result is an 8×8 linear system which is used to iteratively solve for the optimal set of coregistration parameters.

Each time through the loop, the resulting delta updates are added to the current best estimate. Note this means the constants in Table 2 are recomputed each time through the loop. Iteration stops when either the amount by which E_{tot} drops between iterations falls below a preset threshold or the total number of iterations exceeds a maximum allowable upper bound. The Levenberg-Marquardt [Press *et al.*, 1988] method, has been found to be robust in our past single sensor pose work [Beveridge and Riseman, 1992; Beveridge and Riseman, 1994], and it is used here to find the optimal coregistration parameters.

It is trivial to modify this formulation to handle point rather than line segment features in the CCD data. For each point, simply define two orthogonal planes passing through the point and sum the squared point-to-plane distances. This results in a measure which is the squared Euclidian point-to-ray distance between the 3D point and a ray passing through the focal point and the image point.

4.1 Approximating True Sensor Registration

In principle, the information obtained by coregistration in terms of relative sensor position, coupled with the range information present in the LADAR image, permits a pixel by pixel determination of which LADAR pixels correspond to which CCD pixels. This mapping depends upon the 3D geometry of the scene, and likely cannot be represented in terms of an affine transformation between the two image planes. Performing this pixel-by-pixel scene based registration is computationally demanding. It will also generate non-uniform results. Therefore, true registration is approximated with a 2D affine transform between images.

To compute the transformation, selected points are mapped from LADAR back into the CCD image, and a 2D affine transformation is computed which best aligns these pairs of image points. The points selected are those used in the original coregistration and are therefore on the object of interest. Hence, the registration transformation will be most accurate where it matters most, on the object model. The problem of solving for the transformation which minimizes the sum of squared distances between corresponding points subject to a 2D affine transformation is relatively simple. In this particular case, only similarity transformations are considered, but more general affine transformations could certainly be accommodated.

5 Experimental Results

Two sets of experiments are presented. The first is a controlled sensitivity study on synthetic data designed to test the robustness and accuracy of the algorithm. The synthetic data tests approximate the viewing of vehicle

$$\frac{\partial E_{tot}}{\partial \delta \vec{\omega}_{mc}} = w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right) \cdot \delta \vec{\omega} + \hat{N}_{ci} \cdot \Delta \vec{T}_{mc} \right) \left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right)$$

$$+ w_{ml} \sum_{i=1}^{n_l} \lambda_{li} \left(M_{mci}^{2} \delta \vec{\omega} + M_{mci} \Delta \vec{T}_{mc} + M_{mci} \Delta \vec{T}_{cl} \right)$$

$$+ w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\hat{N}_{ci} \cdot \left(\vec{P}_{mij}^{e} + \vec{T}_{mc}^{e} \right) \right) \left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right)$$

$$+ w_{ml} \sum_{i=1}^{n_l} \lambda_{li} M_{mci} \left(\vec{P}_{mi}^{e} + \vec{T}_{mc}^{e} + \vec{T}_{cl}^{e} - \vec{P}_{li} \right) \qquad (7)$$

$$\frac{\partial E_{tot}}{\partial \Delta \vec{T}_{mc}} = w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right) \cdot \delta \vec{\omega} + \hat{N}_{ci} \cdot \Delta \vec{T}_{mc} \right) \hat{N}_{ci}$$

$$+ w_{ml} \sum_{i=1}^{n_l} \lambda_{li} \left(M_{mci} \delta \vec{\omega} + \Delta \vec{T}_{mc} + \Delta \vec{T}_{cl} \right)$$

$$+ w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\hat{N}_{ci} \cdot \left(\vec{P}_{mij}^{e} + \vec{T}_{mc}^{e} \right) \right) \hat{N}_{ci}$$

$$+ w_{ml} \sum_{i=1}^{n_l} \lambda_{li} \left(\vec{P}_{mi}^{e} + \vec{T}_{mc}^{e} + \vec{T}_{cl}^{e} - \vec{P}_{li} \right) \qquad (8)$$

$$\frac{\partial E_{tot}}{\partial \Delta \vec{T}_{cl}} = w_{cl} \sum_{i=1}^{n_l} \lambda_{li} \left(M_{mci} \delta \vec{\omega} + \Delta \vec{T}_{mc} + \Delta \vec{T}_{cl} \right)$$

$$+ w_{cl} \sum_{i=1}^{n_l} \lambda_{li} \left(\vec{P}_{mi}^{e} + \vec{T}_{mc}^{e} + \vec{T}_{cl}^{e} - \vec{P}_{li} \right) \qquad (9)$$

Table 1: Partial derivatives of coregistration error with respect to free variables

sized objects at 500 meters and thus reflect the expected RSTA conditions. The second test is a demonstration on actual data collected at Fort Carson last November. The demonstration illustrates coregistration converging upon both an object pose estimate and sensor-to-sensor coregistration estimate for an actual LADAR and CCD image. Hand picked features on an M60 model are passed as input to the coregistration algorithm.

5.1 Sensitivity Analysis with Synthetic Data

The synthetic data was designed to approximate the viewing of vehicle sized objects at 500 meters. The synthetic CCD sensor has a 4° field of view and generates a 512×512 image; the LADAR images 6 pixels per meter at 500 meters. The sensors are separated by 1 meter. Each model is located 500 meters from sensors along the focal axis of the CCD sensor.

The weights in the coregistration error, λ_{ci}, λ_{li}, w_{mc} and w_{mcl}, are all set to 1.0. The convergence threshold for E_{tot} is 1^{-4} in all tests and the maximum number of iterations is 20.

Two sets of experiments were conducted: I) sensitivity to noise in initial coregistration estimate, and II) sensitivity to noisy image data. Four different models were used. These models exhibit different geometric characteristics including planarity or lack of planarity, symmetry or lack of symmetry, and sparse versus fuller numbers of features.

5.1.1 The Models

Both tests were run on four synthetic models. These models are shown in Figure 3. The exact coordinates of the points comprising the models are given and line segments used for CCD are also shown. The first model, the 'trapezoid', is a planar asymmetric object and tests coregistration performance for coplanar sets of points. The second model, the 'cube', tests coregistration performance for a non-planar symmetric object. The third

model, the 'wedge', is similar to the cube, but is asymmetric. The fourth model, the 'tetrahedron', is a small, non-planar, asymmetric model. The tetrahedron is the model closest in size and structure to features selected on the M60 and used in the real data test in Section 5.4.

5.2 Test I

The goal of test I is to probe the sensitivity of the algorithm to noise in the initial coregistration estimate. The convergence properties of the algorithm are tested both with regard to the number of iterations and the quality of the final solution for a set of noisy initial coregistration estimates. Noise is introduced into the orientation estimate, the translation estimate of the sensors relative to the object model, and the relative translation between the sensors. A baseline error E_{tot} is established for each model by running the algorithm with no error in the initial coregistration estimate.

To generate the test data with known ground truth for the coregistration, both LADAR and CCD features are projected for each model positioned 500 meters from the sensors along the CCD's projection axis. The sensors had no offset relative to each other. Then the sensor positions were perturbed a fixed amount in a random direction. This means they were either rotated, translated, or both, by a fixed amount in a randomly chosen direction. For rotation, both sensors were pertubed together so as to keep the sensor images coplanar.

The results are shown in Table 3. Each row indicates a specific set of experiments. The first column indicates the model used in the test. The second column, labeled ΔR, gives the amount of orientation error (in radians) introduced into the initial coregistration estimate. The third column, labeled ΔT, gives the amount of translation error (in meters) introduced into the initial coregistration estimate. Translation noise is added both to the position of the sensors together relative to the object model and subsequently to the LADAR sensor's posi-

$$A = w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right) \left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right)^{T} + w_{ml} \sum_{i=1}^{n_l} \lambda_{li} M_{mci}^{2} \qquad (11)$$

$$B = w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right) \hat{N}_{ci}^{T} + w_{ml} \sum_{i=1}^{n_l} \lambda_{li} M_{mci} \qquad (12)$$

$$C = w_{ml} \sum_{i=1}^{n_l} \lambda_{li} M_{mci} \qquad (13)$$

$$D = w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\hat{N}_{ci} \left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right)^{T} \right) + w_{ml} \sum_{i=1}^{n_l} \lambda_{li} M_{mci} \qquad (14)$$

$$E = w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \hat{N}_{ci} \hat{N}_{ci}^{T} + w_{ml} \sum_{i=1}^{n_l} \lambda_{li} I_3 \qquad (15)$$

$$F = w_{ml} \sum_{i=1}^{n_l} \lambda_{li} I_3 \qquad (16)$$

$$G = w_{mc} \sum_{i=1}^{n_l} \lambda_{li} M_{mci} \qquad (17)$$

$$H = w_{mc} \sum_{i=1}^{n_l} \lambda_{li} I_3 \qquad (18)$$

$$J = w_{mc} \sum_{i=1}^{n_l} \lambda_{li} I_3 \qquad (19)$$

$$\vec{K} = -w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\hat{N}_{ci} \cdot \left(\vec{P}_{mij}^{e} + \vec{T}_{mc}^{e} \right) \right) \left(\vec{P}_{mij}^{e} \times \hat{N}_{ci} \right)$$
$$\quad - w_{ml} \sum_{i=1}^{n_l} \lambda_{li} M_{mci} \left(\vec{P}_{mi}^{e} + \vec{T}_{mc}^{e} + \vec{T}_{cl}^{e} - \vec{P}_{li} \right) \qquad (20)$$

$$\vec{L} = -w_{mc} \sum_{i=1}^{n_c} \sum_{j=1}^{2} \lambda_{ci} \left(\hat{N}_{ci} \cdot \left(\vec{P}_{mij}^{e} + \vec{T}_{mc}^{e} \right) \right) \hat{N}_{ci} - w_{ml} \sum_{i=1}^{n_l} \lambda_{li} \left(\vec{P}_{mi}^{e} + \vec{T}_{mc}^{e} + \vec{T}_{cl}^{e} - \vec{P}_{li} \right) \quad (21)$$

$$\vec{M} = -w_{ml} \sum_{i=1}^{n_l} \lambda_{li} \left(\vec{P}_{mi}^{e} + \vec{T}_{mc}^{e} + \vec{T}_{cl}^{e} - \vec{P}_{li} \right) \qquad (22)$$

Table 2: Constant matrices and vectors in linear update equation.

tion relative to the CCD. The fourth column, labeled N, gives the number of trials out of 10 which converged successfully. This is defined to be trials for which the final $\overline{E_{tot}}$ is less than 1.0. The fifth column, labeled $\overline{\text{Its.}}$, gives the average number of iterations required to converge. The sixth column, labeled $\overline{E_{tot}}$, gives the average total coregistration error E_{tot} for the N successful trials.

The results indicate the coregistration algorithm really ought not be counted upon to function for ΔR noise greater than 0.5, at which level the algorithm typically succeeds. The only exception is the Tetrahedron, for which 2 out of the 10 trials did not recover the true coregistration for the test in which $\Delta R = 0.5$ and $\Delta T = 0$. Not surprisingly, the algorithm is much more robust with respect to noise in the translation estimate, succeeding for all four models even with $\Delta T = 500$.

5.3 Test II

Test II tests the effects of noisy image data on the accuracy of the recovered coregistration parameters. Varying amounts of noise are added to the image measurements and then the recovered coregistration parameters are compared to the ground truth coregistration. To obtain image data with known ground truth, the four object models are projected into the LADAR and CCD images using the same 500 meter to object range as in Test I.

Gaussian noise is added to the image points. The noise is independently applied to the X and Y image axes. The coregistration algorithm is then run using the true coregistration parameters for the initial estimate. The results for all these test are summarized in Table 4. Results for each of the models and seven different image noise combinations are shown. The first two columns indicate the σ for the Gaussian noise added to the CCD image and

LADAR image points respectively. The third and fourth columns report the average and standard deviation for the error (in radians) between the true and recovered model to CCD orientation estimate R_{mc}. The averages are for 10 trials. The fifth and sixth columns report the average and standard deviation for the error in model to CCD translation estimate. The seventh and eighth columns report the average and standard deviation for the error in the LADAR to CCD translation estimate. Translations are given in meters.

Errors in LADAR produce more error in the coregistration estimate than do comparable errors in the CCD data. There are two reasons for this. First, the pixel resolution is lower for LADAR in these simulations, and hence a one pixel σ represents greater uncertainty in the scene for LADAR than it does for the CCD. Second, the Euclidean distance measured in the LADAR-to-model error exerts more pull over the final coregistration than does the point-to-plane distance associated with the CCD-to-model error.

The coregistration algorithm is recovering the object pose and sensor registration to within 2% of the total sensor to 500 meter object-to-sensor distance of image noise with $\sigma = 1$. This is true for all four object models tested. The average error in the translation of the model relative to the sensors, $\vec{T}_{mc}$, is 7, 5, 4 and 6 meters respectively for the Trapezoid, Cube, Wedge, and Tetrahedron. This indicates translation error is somewhat, but not overly, dependent upon specific model geometry, and the roughly 5 meters of error out of 500 will support the types of coregistration matching for which this algorithm is being developed.

The average error in orientation for $\sigma = 1$ in both CCD and LADAR ranges from 0.008 to 0.018 radians, and indicates the coregistration orients the model to the
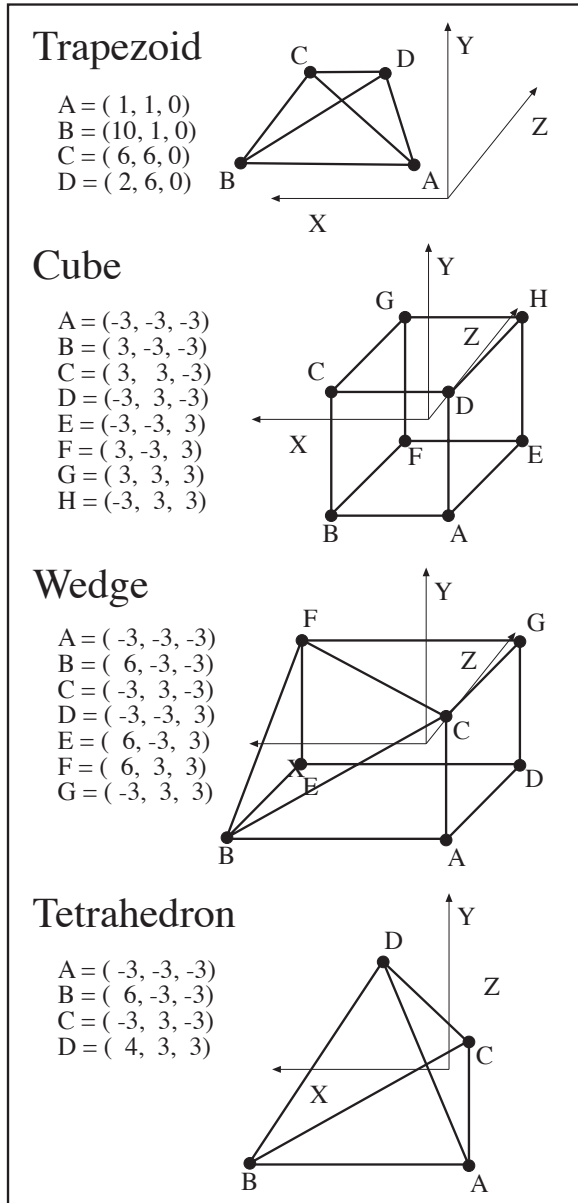
Figure 3: Four models used to test coregistration code.

Model	ΔR	ΔT	N	$\overline{\text{Its.}}$	$\overline{E_{tot}}$
Trapeziod	0	0	1	-	7.13e-10
"	0	40	10	3.0	7.12e-10
"	0.50	0	10	3.9	7.10e-10
"	0.25	20	10	3.0	7.12e-10
"	0.50	40	10	3.0	7.12e-10
"	0.90	100	9	4.9	7.28e-10
"	3.14	0	0	-	-
"	0	500	10	3.5	7.41e-10
Cube	0	0	1	-	5.70e-10
"	0	40	10	3.0	5.59e-10
"	0.50	0	10	3.9	5.84e-10
"	0.25	20	10	3.0	5.59e-10
"	0.50	40	10	3.0	5.59e-10
"	0.90	100	10	5.5	5.61e-10
"	3.14	0	8	11.9	5.61e-10
"	0	500	10	3.0	5.59e-10
Wedge	0	0	1	-	0.62e-9
"	0	40	10	3.0	9.04e-9
"	0.50	0	10	6.3	3.35e-1
"	0.25	20	10	3.0	9.04e-9
"	0.50	40	10	3.0	9.04e-9
"	0.90	100	7	6.6	4.75e-1
"	3.14	0	3	9.7	2.12e-1
"	0	500	10	4.3	0.62e-1
Tetrahedron	0	0	1	-	3.74e-10
"	0	40	10	3.0	5.42e-10
"	0.50	0	8	3.8	5.78e-10
"	0.25	20	10	3.0	5.42e-10
"	0.50	40	10	3.0	5.42e-10
"	0.90	100	8	4.9	5.44e-10
"	3.14	0	1	6.0	5.42e-10
"	0	500	10	3.6	5.42e-10

Table 3: Test I Results

sensors to within 1 degree. Increasing σ to 5.0 degrades performance considerably, with error ranging from 0.047 to 0.086 radians, and thus introducing up to 5 degrees of orientation error.

The model to CCD orientation and translation estimates are coupled, since introducing rotation may introduce associated translations. It is therefore not surprising to see large model-to-CCD translation errors associated with large orientation errors. In contrast, the LADAR-to-CCD translation does not experience this coupling because the LADAR and CCD are not permitted to rotate relative to each other. This explains the much lower average errors in LADAR-to-CCD translations, $\vec{T}_{cl}$, given in the last two columns of Table 4

5.4 Demonstration on Real Data

Four corresponding points on a 3D M60 object model, in the Fort Carson nov31553c color image, and in the nov31205l LADAR image were hand picked using the Rangeview program described elsewhere in these proceedings [Goss et al., 1994]. The four pairs of corresponding points were used to generate six corresponding pairs of line segments for the model-to-CCD error term. These corresponding features, intrinsic camera parameters, and an initial coregistration estimate were passed to the coregistration algorithm.

The algorithm converged to a solution in 6 iterations. The object-to-sensor pose and CCD-to-LADAR registration for the initial coregistration estimate and after the first coregistration iteration are shown in Figure 4. The initial coregistration estimate, shown in Figure 4a, places the M60 roughly 50 meters too high in the scene. The CCD to LADAR sensor registration is off by nearly 15 pixels. Using the LADAR hypothesis generation techniques described in [Beveridge et al., 1994a], we expect much better initial coregistration estimates.

Even with this poor initial estimate, the coregistration recovers what appears to be the proper coregistration [2].

[2] We can only say appears, since precise ground truth is

7

Figure 4: Illustrating the coregistration algorithm running on real data. a) initial errorful coregistration estimate, b) improvement after first iteration of algorithm.

Most of the visibly evident correction is made in the very first iteration. The result of this first iteration is shown in Figure 4b. The M60 moves to near its proper true position relative to the LADAR data, and the CCD data shifts up relative to the LADAR to near its proper position. The light-to-dark skyline is now within several pixels of the proper position in the upper right hand corner of the image. In contrast, in Figure 4a this horizon line is down at the level of the M60 in the LADAR.

6 Conclusion

The coregistration algorithm presented here is quite new, and more tests are planned for both controlled synthetic and real data. The early results are encouraging, and future efforts will push development in two directions. First, the construction of new multisensor matching techniques which utilize this particular eight degree

---

not available for this data.

of freedom coregistration algorithm. Second, the development of alternative coregistration procedures which assume different numbers of sensors and different relative constraints between sensors.

A case of particular interest derives from a scenario in which multiple vehicles using GPS are performing a coordinated RSTA function. This may be thought of as a multisensor problem in which relative position of sensors is well constrained while relative orientation is not. This follows from the fact that GPS recovers position quite accurately, while determining precise vehicle orientation, to within fractions of a degree, is more difficult. We have begun to formulate a coregistration algorithm for this multiple vehicle problem.

References

[Aggarwal, 1990] J. K. Aggarwal. Multisensor Fusion for Automatic Scene Interpretation. In Ramesh C. Jain

Noise σ		R_{mc}		T_{mc}		T_{cl}	
C	L	Avg.	Sd.	Avg.	Sd.	Avg.	Sd.
Trapezoid							
0.5	0.5	0.009	0.006	4	3	0.05	0.03
1	1	0.018	0.013	7	6	0.09	0.07
5	0	0.044	0.031	1	1	0.20	0.13
0	5	0.092	0.064	40	31	0.48	0.32
5	5	0.086	0.052	33	23	0.48	0.33
20	0	0.183	0.082	4	4	0.81	0.47
0	20	0.418	0.313	188	151	1.86	1.25
Cube							
0.5	0.5	0.006	0.005	3	2	0.01	0.01
1	1	0.010	0.007	5	3	0.01	0.01
5	0	0.011	0.012	3	4	0.02	0.05
0	5	0.070	0.028	34	14	0.12	0.10
5	5	0.048	0.031	21	15	0.05	0.04
20	0	0.034	0.040	10	15	0.06	0.18
0	20	0.318	0.190	155	92	0.62	0.74
Wedge							
0.5	0.5	0.005	0.004	2	2	0.01	0.02
1	1	0.008	0.007	4	3	0.02	0.03
5	0	0.007	0.012	3	5	0.01	0.02
0	5	0.055	0.031	28	15	0.06	0.06
5	5	0.047	0.039	22	18	0.12	0.16
20	0	0.029	0.042	9	16	0.01	0.01
0	20	0.213	0.095	106	48	0.27	0.25
Tetrahedron							
0.5	0.5	0.005	0.004	2	2	0.01	0.02
1	1	0.010	0.007	6	4	0.03	0.03
5	0	0.016	0.018	7	7	0.05	0.11
0	5	0.077	0.060	37	30	0.27	0.38
5	5	0.063	0.057	30	30	0.17	0.24
20	0	0.074	0.091	25	31	0.24	0.49
0	20	0.209	0.111	97	50	1.15	1.53

Table 4: Test II Results.

and Anil K. Jain, editors, *Analysis and Interpretation of Range Images*, chapter 8. Springer-Verlag, 1990.

[Beveridge and Riseman, 1992] J. Ross Beveridge and Edward M. Riseman. Hybrid Weak-Perspective and Full-Perspective Matching. In *Proceedings: IEEE 1992 Computer Society Conference on Computer Vision and Pattern Recognition*, pages 432 – 438. IEEE Computer Society, June 1992.

[Beveridge and Riseman, 1994] J. Ross Beveridge and Edward M. Riseman. Optimal Geometric Model Matching Under Full 3D Perspective. In *Second CAD-Based Vision Workshop*, pages 54 – 63. IEEE Computer Society Press, February 1994. (Submitted to CVGIP-IU).

[Beveridge et al., 1990] J. Ross Beveridge, Rich Weiss, and Edward M. Riseman. Combinatorial Optimization Applied to Variable Scale 2D Model Matching. In *Proceedings of the IEEE International Conference on Pattern Recognition 1990, Atlantic City*, pages 18 – 23. IEEE, June 1990.

[Beveridge et al., 1991] J. Ross Beveridge, Rich Weiss, and Edward M. Riseman. Optimization of 2-Dimensional Model Matching. In Hatem Nasr, edi-

tor, *Selected Papers on Automatic Object Recognition (originally appeared in DARPA Image Understanding Workshop, 1989)*, SPIE Milestone Series. SPIE, Bellingham, WA, 1991.

[Beveridge et al., 1994a] J. Ross Beveridge, Allen Hanson, and Durga Panda. RSTA Research of the Colorado State, University of Massachusetts and Alliant Techsystems Team. In *Proceedings: Image Understanding Workshop*, page (to appear). Morgan Kaufmann, November 1994.

[Beveridge et al., 1994b] J. Ross Beveridge, Durga P. Panda, and Theodore Yachik. November 1993 Fort Carson RSTA Data Collection Final Report. Technical Report CSS-94-118, Colorado State University, Fort Collins, CO, January 1994.

[Beveridge, 1993] J. Ross Beveridge. *Local Search Algorithms for Geometric Obejct Recognition: Optimal Correspondence and Pose*. PhD thesis, University of Massachuesetts at Amherst, May 1993.

[Bolles et al., 1993] Robert C. Bolles, H. Harlyn Baker, and Marsha Jo Hannah. The JISCT Stereo Evaluation. In *Proceedings: Image Understanding Workshop*, Los Altos, CA, April 1993. ARPA, Morgan Kaufmann.

[Eason and Gonzalez, 1992] R. O. Eason and R. C. Gonzalez. Least-Squares Fusion of Multisensory Data. In Mongi A. Abidi and Rafael C. Gonzalez, editors, *Data Fusion in Robotics and Machine Intelligence*, chapter 9. Academic Press, 1992.

[Goss et al., 1994] Michael E. Goss, J. Ross Beveridge, Mark Stevens, and Aaron Fuegi. Visualization and Verification of Automatic Target Recognition Results Using Combined Range and Optical Imagery. In *Proceedings: Image Understanding Workshop*, page (to appear), Los Altos, CA, November 1994. ARPA, Morgan Kaufmann.

[Grimson, 1990] W. Eric L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints*. MIT Press, Cambridge, MA, 1990.

[Hebert et al., 1990] Martial Hebert, Takeo Kanade, and InSo Kweon. 3-D Vision Techniques for Autonomous Vehicles. In Ramesh C. Jain and Anil K. Jain, editors, *Analysis and Interpretation of Range Images*, chapter 7. Springer-Verlag, 1990.

[Hel-Or and Werman, 1993] Y. Hel-Or and M. Werman. Absolute Orientation from Uncertain Data: A Unified Approach. In *Proceedings: Computer Vision and Pattern Recognition*, pages 77 – 82. IEEE Computer Society Press, June 1993.

[Huttenlocher and Ullman, 1990] Daniel P. Huttenlocher and Shimon Ullman. Recognizing solid objects by alignment with an image. *Internation Journal of Computer Vision*, 5(2):195 – 212, November 1990.

[Kumar and Hanson, 1994] Rakesh Kumar and Allen R. Hanson. Robust Methods for Estimating Pose and a Sensitivity Analysis. *CVGIP:Image Understanding*, 11:(to appear in November), 1994.

[Kumar, 1989] Rakesh Kumar. Determination of Camera Location and Orientation. In *Proceedings: Image*

Understanding Workshop, pages 870 – 881, Los Altos, CA, June 1989. DARPA, Morgan Kaufmann Publishers, Inc.

[Kutulakos and Dyer, 1994] Kiriakos N. Kutulakos and Charles R. Dyer. Occluding Contour Detection Using Affine Invariants and Purposive Viewpoint Control. In *Proceedings: Computer Vision and Pattern Recognition*, pages 323 – 330. IEEE Computer Society Press, June 1994.

[Lowe and Binford, 1985] David G. Lowe and Thomas O. Binford. The Recovery of Three-Dimensional Structure from Image Curves. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(3):320 – 325, 1985.

[Lowe, 1991] David G. Lowe. Fitting Parameterized Three-Dimensional Models to Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(5):441 – 450, May 1991.

[Magee *et al.*, 1985] M. J. Magee, B. A. Boyter, C. H. Chien, and J. K. Aggarwal. Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(6):629 – 637, November 1985.

[Press *et al.*, 1988] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.

[Stentz and Goto, 1987] A. Stentz and Y. Goto. The CMU Navigational Architecture. In *Proceedings: Image Understanding Workshop*, pages 440–446, Los Angeles, CA, February 1987. ARPA, Morgan Kaufmann.

[Y. Hel-Or and M. Werman, 1994] Y. Hel-Or and M. Werman. Constraint-Fusion for Interpretation of Articulated Objects. In *Proceedings: Computer Vision and Pattern Recognition*, pages 39 – 45. IEEE Computer Society Press, June 1994.