

Coregistration of Range and Optical Images Using Coplanarity and Orientation Constraints *

Anthony N. A. Schwickerath

J. Ross Beveridge

Computer Science Department
Colorado State University
Fort Collins, Colorado 80523
schwicke/ross@cs.colostate.edu

Abstract

A least-squares method simultaneously solves for the model-to-sensor-suite pose and sensor-to-sensor registration. The development is for a sensor-suite containing separate range and optical sensors. To address outliers and, more generally, match finding, a statistical method (median filtering) and a search method (local search) are developed. Sensitivity to Gaussian noise and the choice of initial pose estimates is investigated on synthetic data. Both of the matching methods are demonstrated on real data.

1 Introduction

Coregistration describes a process which simultaneously refines both the estimated 3D pose of an object relative to a sensor suite as well as the registration parameters relating the coordinate systems of a range sensor and an optical sensor. It extends single sensor pose work [14, 17, 9, 19] by imposing constraints on both sensor and object geometry.

Coregistration will support multi-sensor object verification but is not itself an object recognition tool. It presupposes a prior system has hypothesized the presence of specific objects and provided approximate positions and orientations relative to the sensor suite. Others have already undertaken this hypothesis generation work [7, 29, 8, 16].

As in previous pose determination work [9, 19], the best pose+registration parameters jointly minimize a quadratic, non-linear function. An iterative technique solves for the 6 pose and 2 registration parameters which jointly minimize a weighted sum of Kumar's [19] measure for optical features and squared 3D Euclidean distance for range data.

Least-squares techniques are sensitive to outliers. Median-filtering more robustly estimates the true coregistration and is compared with a local search matching capability [3, 4]. Results are presented for both controlled synthetic and real data.

2 Background & Motivation

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. Lowe [20], Huttenlocher [15], Grimson [10], and we [5, 6, 3, 4] have all proposed system within this paradigm. Our work here provides an example of extending these methods to multiple constrained sensors.

Others have worked on problems similar to the coregistration problem discussed in this paper. Herbert [11] presents a least-squares mechanism for computing the 2D registration between a range and color CCD sensor based upon corresponding image points in the two sensor images. Both Eason [9] and Hel-Or [13] develop least-squares multisensor pose algorithms. However, they do not support simultaneous sensor registration. In terms of constraints, all of these methods assume a known sensor-to-sensor registration.

More recent work by Hel-Or and Werman [30, 12] adds degrees of freedom to account for articulated objects and handles variable constraints in a single extended Kalman filter formulation. Their general Kalman filter approach could be applied to the coregistration problem formulated here. However, For reasons of efficiency we have intentionally used what Hel-Or and Werman call the 'parametric' approach. Future work will test the relative merits of the 'parametric' versus Kalman filter approaches.

Coregistration is a form of model-based sensor fu-

*This work was sponsored by the Advanced Research Projects Agency (ARPA) under grant DAAH04-93-G-422, monitored by the U. S. Army Research Office.

sion. In summarizing past sensor fusion work, Aggarwal [1] 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 [21, 27]. He further states that to properly perform mixed-modality sensor fusion, **coordinate transformations between images need to be adaptively determined**. Coregistration, as presented in this paper, is a working example of such an adaptive process.

3 The Coregistration Algorithm

The coregistration procedure derived here assumes known 3D model features, an optical sensor modeled as a pin-hole camera, and a range sensor. Thus, as input, it takes corresponding line segment features on the model and in the optical image, corresponding point features on the model and in the range image, and an initial coregistration estimate.

In our experiments, the 3D object model features are derived from BRL-CAD models of military vehicles. Features appropriate for matching to range as well as separate features appropriate for matching to optical imagery are extracted from the CAD models. The range pixels in the LADAR images serve as range point features. For the CCD data, a steerable filter and a global edge fitting procedure is used to find locally optimal straight edge segments [22]. How these features are derived is further explained in [28, 22, 2].

A coplanarity constraint limits the freedom of movement of the range sensor relative to the optical sensor. Thus, the range reference coordinate system may translate in the common x-y image plane of the two sensors, but otherwise the two sensors move together. In general, this is very restrictive assumption. However, it is adequate and even desirable for our application domain. Coregistration is being developed to recognize military vehicles at a distance. For pixels on such a vehicle, pixel-to-pixel movement between the two images resulting from rotation is essentially equivalent to that induced by translation. Expressing all image-to-image movement as planar translation limits the overall non-linearity of the coregistration error function.

3.1 The Coregistration Equations

The best coregistration estimate minimizes a quadratic error of fit between corresponding object model and sensor features.

$$E_{fit} = \alpha_{fit} E_{fit,o} + (1 - \alpha_{fit}) E_{fit,r} \quad (1)$$

The constituent parts of E_{fit} are illustrated in Figure 1. The first term, $E_{fit,o}$, measures distance be-

tween corresponding optical and model features. This term is precisely the point-to-plane error criterion defined by Kumar [17, 19] for computing camera-to-model pose¹. The second, $E_{fit,r}$, is simply the sum-of-squared Euclidean distances between corresponding model and range points.

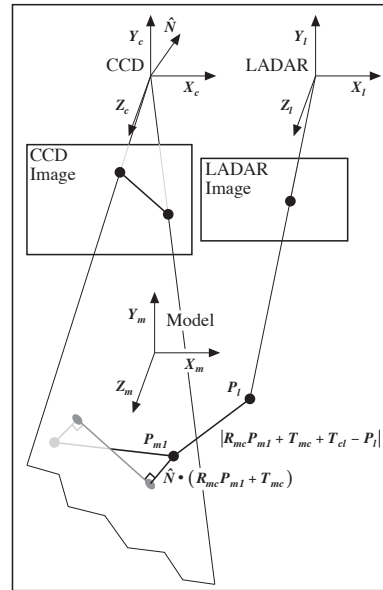


Figure 1: Illustrating distance errors which define optimal coregistration.

The weighting term $0 \leq \alpha_{fit} \leq 1$ controls the relative importance of the optical and the range data. The terms $E_{fit,o}$ and $E_{fit,r}$ are normalized between $[0, 1]$ based upon the expected amount of noise present in the features², and consequently E_{fit} also falls in this range. This normalization allows comparison of data from two separate sources.

The model-to-optical fit error, $E_{fit,o}$, measures the distance between endpoints of 3D line segments on the object model and 3D planes defined by corresponding line segments found in the optical image. These planes are defined by three points: the two endpoints of the optical line segment and the focal point of the optical sensor. If we have n_o model-to-optical correspondences, we write the model-to-optical fit error as:

$$E_{fit,o} = w_{fit,o} \sum_{i=1}^{n_o} \sum_{j=1}^2 \lambda_{oi} \left(\hat{N}_{oi} \cdot \left(R_{mo} \vec{P}_{mij} + \vec{T}_{mo} \right) \right)^2 \quad (2)$$

¹After developing this measure, Kumar developed others which are more robust but which also require additional normalization.

²While a rigorous and complete noise model is not developed, the Gaussian noise assumption underlies least-squares.

$$w_{fit,o} = \frac{1}{2n_o\tau_{m_o}^2 \sum_i \lambda_{oi}} \quad (3)$$

where \hat{N}_{oi} is the normal to the plane defined by the i th image line, R_{m_o} is the rotation from model to optical sensor coordinates, \vec{P}_{mij} is the j th endpoint of the i th model line segment in model coordinates, and \vec{T}_{m_o} is the translation from model to optical sensor coordinates. The weighting term λ_{oi} is typically 1, but can be used to bias some features over others. For example, λ_{oi} can be used to weight lines based on inverse distance to perform a normalization similar to that considered in Kumar's later measures [17, 19].

The τ_{m_o} in the normalizing weight for optical fit, $w_{fit,o}$, represents the maximum anticipated error. Given the Gaussian noise assumption, this threshold can be thought of in terms of the standard deviation, since only 4% of the data will fall more than 2 standard deviations from the true value. While this assumption does not absolutely place $E_{fit,o}$ in the range $[0, 1]$, this will generally be the case.

For range data, 3D Cartesian points are formed by back-projecting range pixels into the scene. The model-to-range fit error, $E_{fit,r}$, is defined to be the squared Euclidean distance between each back-projected range point and its corresponding model point. We write this as:

$$E_{fit,r} = w_{fit,r} \sum_{i=1}^{n_r} \lambda_{ri} \left| \left(\left(R_{or} \vec{P}_{m_{oi}} + \vec{T}_{or} \right) - \vec{P}_{ri} \right) \right|^2 \quad (4)$$

$$w_{fit,r} = \frac{1}{n_r \tau_{m_r}^2 \sum_i \lambda_{ri}} \quad (5)$$

$$\vec{P}_{m_{oi}} = R_{m_o} \vec{P}_{mi} + \vec{T}_{m_o} \quad (6)$$

where $\vec{P}_{m_{oi}}$ is the i th model point mapped into optical sensor coordinates, \vec{P}_{ri} is the corresponding i th measured range point, and \vec{T}_{or} (which is of the form $(t_{ol_x}, t_{ol_y}, 0)$) is the optical-to-range sensor registration parameter with the coplanarity constraint built in. The lack of a rotation parameter between mc mapped points and the range sensor coordinate system constrains the sensor-to-sensor orientation. Notice that these constraints retain the same degree of nonlinearity (degree 2) found in the original Kumar [17, 19] and Horn [14] equations. This is desirable, since increasing the nonlinearity of a system of equations tends to increase the instability of the solution. Again, we use a weighted form based on a threshold τ_{m_r} , which keeps us generally in the range $[0, 1]$.

If we treat R_{m_o} as a 3×3 rotation matrix, solving for R_{m_o} by minimizing Equation 1 and allowing all 9 terms to vary independently violates the constraint

that R_{m_o} be a rotation matrix. While the matrix terms $(r_{mc_{1,1}} \dots r_{mc_{3,3}})$ could be constructed in such a way as to allow only rigid rotations, this would increase the degree of nonlinearity in the equation. Kumar [17, 19] suggests a better approach: Rodriguez's formula, which is an approximation appropriate for small rotations. To rotate a point \vec{P}_{mi} by an amount R_{m_o} , we write

$$\begin{aligned} R_{m_o} \vec{P}_{mi} &= R_{m_o}^e \vec{P}_{mi} + \delta \vec{\omega}_{m_o} \times \left(R_{m_o}^e \vec{P}_{mi} \right) \\ &= \vec{P}_{mi}^e + \delta \vec{\omega}_{m_o} \times \vec{P}_{mi}^e \end{aligned} \quad (7)$$

(Note: to provide a more compact notation the vector \vec{P}_{mi}^e is introduced as the current estimate of the transformed model point.) Here, $R_{m_o}^e$ is the current estimated rotation (3×3) matrix and $\delta \vec{\omega}_{m_o}$ is the small rotation update represented as a unit rotational axis scaled by the rotational magnitude.

In order to minimize E_{fit} in equation 1 with respect to $\delta \vec{\omega}_{m_o}$, $\Delta \vec{T}_{m_o}$ and $\Delta \vec{T}_{or}$, the partial derivatives with respect to each are set to zero and the resulting system of equations solved for the coregistration update parameters. These partial derivatives are shown in [26]. In these equations, we simplify these by rewriting the weighting terms as $w_{m_o} = \alpha_{fit} w_{fit,o}$ and $w_{m_r} = (1 - \alpha_{fit}) w_{fit,r}$.

The following matrix $M_{m_{oi}}$ is introduced to simplify the expressions.

$$M_{m_{oi}} = \begin{pmatrix} 0 & V_{m_{oi},z} & -V_{m_{oi},y} \\ -V_{m_{oi},z} & 0 & V_{m_{oi},x} \\ V_{m_{oi},y} & -V_{m_{oi},x} & 0 \end{pmatrix} \quad (8)$$

$$V_{m_{oi}} = R_{or} \vec{P}_{mi}^e \quad (9)$$

The matrix $M_{m_{oi}}$ is the vector product with $R_{or} \vec{P}_{mi}^e$. In other words: $M_{m_{oi}} \vec{V} = \vec{V} \times R_{or} \vec{P}_{mi}^e$. It is also the partial derivative of $\delta \vec{\omega}_{m_o} \times R_{or} \vec{P}_{mi}^e$ with respect to $\delta \vec{\omega}_{m_o}$.

Setting the partial derivatives to zero yields 9 linear equations in 9 unknowns. These linear equations may be written as:

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

where the constants A through M are defined in Table 1. Since Δ_{or} has the form $(\delta t_{ol_x}, \delta t_{ol_y}, 0)$, we can drop the rightmost column and bottom row of the 9×9 matrix in equation 8. The result is an 8×8 linear system which is used to iteratively solve for the optimal set of pose+registration parameters.

$$A = w_{mo} \sum_{i=1}^{n_o} \sum_{j=1}^2 \lambda_{oi} \left(\vec{P}_{mij}^e \times \hat{N}_{oi} \right) \left(\vec{P}_{mij}^e \times \hat{N}_{oi} \right)^T + w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} M_{moi}^2 \quad (11)$$

$$B = w_{mo} \sum_{i=1}^{n_o} \sum_{j=1}^2 \lambda_{oi} \left(\vec{P}_{mij}^e \times \hat{N}_{oi} \right) \hat{N}_{oi}^T + w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} M_{moi} \quad (12)$$

$$C = w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} M_{moi} \quad (13)$$

$$D = w_{mo} \sum_{i=1}^{n_o} \sum_{j=1}^2 \lambda_{oi} \left(\hat{N}_{oi} \left(\vec{P}_{mij}^e \times \hat{N}_{oi} \right)^T \right) + w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} M_{moi} \quad (14)$$

$$E = w_{mo} \sum_{i=1}^{n_o} \sum_{j=1}^2 \lambda_{oi} \hat{N}_{oi} \hat{N}_{oi}^T + w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} I_3 \quad (15)$$

$$F = w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} I_3 \quad (16)$$

$$G = w_{mo} \sum_{i=1}^{n_o} \lambda_{ri} M_{moi} \quad (17)$$

$$H = w_{mo} \sum_{i=1}^{n_o} \lambda_{ri} I_3 \quad (18)$$

$$J = w_{mo} \sum_{i=1}^{n_o} \lambda_{ri} I_3 \quad (19)$$

$$\begin{aligned} \vec{K} &= -w_{mo} \sum_{i=1}^{n_o} \sum_{j=1}^2 \lambda_{oi} \left(\hat{N}_{oi} \cdot \left(\vec{P}_{mij}^e + \vec{T}_{mo}^e \right) \right) \left(\vec{P}_{mij}^e \times \hat{N}_{oi} \right) \\ &\quad - w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} M_{moi} \left(R_{or} \vec{P}_{mi}^e + R_{or} \vec{T}_{mo}^e + \vec{T}_{or}^e - \vec{P}_{ri} \right) \end{aligned} \quad (20)$$

$$\begin{aligned} \vec{L} &= -w_{mo} \sum_{i=1}^{n_o} \sum_{j=1}^2 \lambda_{oi} \left(\hat{N}_{oi} \cdot \left(\vec{P}_{mij}^e + \vec{T}_{mo}^e \right) \right) \hat{N}_{oi} \\ &\quad - w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} \left(R_{or} \vec{P}_{mi}^e + R_{or} \vec{T}_{mo}^e + \vec{T}_{or}^e - \vec{P}_{ri} \right) \end{aligned} \quad (21)$$

$$\vec{M} = -w_{mr} \sum_{i=1}^{n_r} \lambda_{ri} \left(R_{or} \vec{P}_{mi}^e + R_{or} \vec{T}_{mo}^e + \vec{T}_{or}^e - \vec{P}_{ri} \right) \quad (22)$$

Table 1: Constant matrices and vectors in linear update equation.

Each time through the loop, the resulting delta updates ($\delta\omega_{mo}$, ΔT_{mo} , and ΔT_{or}) are added to the current estimate (R_{mo}^e , T_{mo}^e , and T_{or}^e). The constants in Table 1 are recomputed each time through the loop. The algorithm converges when the amount by which E_{fit} drops between successive iterations falls below a preset threshold. Unsuccessful termination occurs if the total number of iterations exceeds a maximum number of iterations. The Levenberg-Marquardt [24] method has been found to be robust in our past single sensor pose work [3], and it is used here to find the optimal pose+registration parameters.

4 Building Matches: Median Filtering and Local Search

The coregistration method above requires a matching between model and image features. As a least-squares method, it requires that this matching be free of outliers. Below, two methods for constructing outlier-free matches are presented.

4.1 Median Filtering

Median filtering [25] handles outliers by fitting to the subset of the data which minimizes the ensemble median error value. It is a robust statistic when

there are less than 50% outliers. This is in contrast to the mean around which least-squares algorithms are based, where a single outlier can radically shift the result. The subset which minimizes the median error must contain no outliers, otherwise it would skew the error, increasing the median.

The space of subsets is combinatoric and hence typically large. To avoid exhaustive search, the space is randomly sampled. Given sufficient samples, the probability of seeing at least one outlier free subset is very high. This yields the optimal fit, and allows us to throw out all data not accounted for by the Gaussian assumption (i.e., outside of two standard deviations of the best fit function, since this will contain 98% of the data effected by Gaussian noise).

The subsets need to be at least large enough to cover the degrees of freedom, so at least 3 optical lines and 1 range point are needed. However, Kumar [18] found that selecting a minimal number of features caused the solution to be sensitive to the Gaussian noise that we assume is overlaid onto the true data. As a consequence, it is better to select a larger subset to stabilize the optimal pose against noise. If we select too large a subset size, however, we greatly reduce

our chances of selecting a subset with no outliers. A compromise must be made between probability and stability.

Once we have minimized the error, we need to select a cutoff point, above which we will consider correspondences to be outliers. We can achieve this either by selecting some *a priori* threshold or by computing one based upon the median. We choose the later method. Assuming a normal distribution, we can set $\text{cutoff} = (a \times s)^2$ where $s = \frac{\min \bar{E}_{fit}}{0.6745}$ is an approximation of the standard deviation for a Gaussian distribution based upon the interquartile range. Setting a to 2.0 filters out data which lies more than two standard deviations above the error, so that the majority of the Gaussian data will be retained.

4.2 Searching Correspondence Space

The process of generating the best correspondence set can be viewed as a combinatorial optimization problem. The brute force expansion of all possible combinations of correspondences, while robust, is also combinatorially explosive. Applying an alternative such as local search [23] will find a *locally* optimal solution. While this makes no guarantees of globality, we can gain confidence in the globality of the optima through multiple search restarts.

So far we have defined a measure of how well the model fits the data, given a set of pose+registration parameters and a correspondence. Using only this measure, we would soon find that the optimal correspondence consisted of a small number of correspondences with zero total fit error. As a result, we also need to consider another error term which offers a penalty for not explaining model features. As in our previous work [3], we combine the omission error (E_{om}) with the fit error to form a total match error per sensor of

$$E_{match,sensor} = E_{fit,sensor} + E_{om,sensor} \quad (23)$$

and the total match error for the suite of

$$E_{match} = \alpha_{match} E_{match,o} + (1 - \alpha_{match}) E_{match,r} \quad (24)$$

Note here that α_{match} is the same as α_{fit} ; it is simply the proportional importance of each sensor.

We consider omission to be the process by which some features or portions of features present in the model are not observed in the image. In the case of range data, the omission will be a function of p , the percentage of points in the model which are not represented in the image. We expect some features to be omitted, but large numbers of omissions should give

cause for alarm. To embody this intuition, an exponential function is used.

$$E_{om,r} = \left\{ \begin{array}{ll} \frac{e^{\alpha_r p} - 1}{e^{\alpha_r} - 1} & \alpha_r \neq 0 \\ p & \alpha_r = 0 \end{array} \right\} \quad (25)$$

The attenuation parameter a_r controls the curvature of the exponential function through α_r ,

$$\alpha_r = 2 \ln \left(\frac{2}{a_r} - 1 \right) \quad (26)$$

Similarly, the omission error for the optical sensor can be thought of in terms of percent of omitted features. Here, however, features can be partially omitted. By applying the exponential error weighting function for each line, the exclusion of complete lines is penalized significantly more than the partial absence of many lines. Consider, for example, the case where we have a square as the model. We expect an image with 75% of each of these four lines to have a lower omission measure than with 100% of each of three lines, even though the total line coverage is equal. We can state a line omission error which embodies these requirements

$$E_{om,o} = \frac{1}{|M_o|} \sum_{i \in} \left\{ \begin{array}{ll} \frac{e^{\alpha_o p_i} - 1}{e^{\alpha_o} - 1} & \alpha_o \neq 0 \\ p_i & \alpha_o = 0 \end{array} \right\} \quad (27)$$

Five parameters specify this match error: τ_o and τ_r for controls the expected noise threshold, a_o and a_r for attenuates the omission response, and α_{match} adjusts for the importance of each sensor.

In local search, the state space is traversed using a series of moves. In our case, the state space is the space of all correspondences ($2^{|M| \times |D|}$ states).

An initial state is generated based upon the hypothesized pose+registration parameters, projecting the model into range and optical sensors. The initial correspondence contains all pairs in which the model and image feature are within a given threshold.

Given this initial state, the set of moves for traversing the search space to an optimal solution is defined by the removal of single correspondences. This simple minded operator allows us to thin out our correspondence, removing outlier data. This also has the advantage that, as the algorithm runs, the neighborhood shrinks, shortening the time required to select the next move. While other, more complete neighborhoods, such as 1- and 2- Hamming distance, have been proposed, this satisfies the tractability requirements of a first pass. The steepest descent transition is made after every possible move from the current state has been evaluated, as in [3]. When no neighbors of the current state offer improvement, a local optima has been reached and the search terminates.

5 Empirical Results

Two empirical studies 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-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, Colorado. The real demonstration illustrates coregistration converging upon both a model-to-sensors pose estimate and sensor-to-sensor coregistration estimate for an actual LADAR and CCD image, using both median filtering and local search matching.

5.1 Synthetic Data

The synthetic optical sensor has a 4° field of view and generates a 512×512 image; the range images, 6 pixels per meter at 500 meters. The sensors are separated by 1 meter. Each model is located 500 meters from the sensors along the focal axis of the optical sensor. The ground truth image data for these tests is obtained for each sensor by synthetically projecting the appropriate model features (lines for optical, points for range) onto the sensor image plane.

Algorithm tuning parameters such as error weighting terms and convergence criteria are constant throughout both experiments. The weights in the coregistration error, λ_{oi} , λ_{ri} , w_{mo} and w_{mr} ³, are all set to 1.0. The convergence threshold for E_{fit} is 10^{-4} . 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. Both tests were run on four synthetic models. The models exhibit different geometric characteristics including planarity or lack of planarity, symmetry or lack of symmetry, and few versus many features. Complete results of these tests are reported in [26].

In Test I, we found that, given perfect image data, the coregistration algorithm could reliably recover correct pose+registration given up to a 30° error in orientation. The correct solution was often found even orientation errors as large as 50° and initial translation errors up to 100 meters. This suggests that, given good data, the algorithm reliably converges upon the optimal set of pose+registration parameters. Test II shows that, given modest image noise ($\sigma = 1$ for both sensors), the final rotation error was within 1° of the correct value. With significantly higher errors, though, ($\sigma = 5$ for both sensors), coregistration yielded a final rotation error around 5° .

³The weights are the combined threshold and α_{fit} term described in [26]

5.2 Real Data

In our previous work[26], instabilities and pathological behavior were found when running coregistration on hand-picked features. This behavior has been traced to outliers present in the hand-picked data. To address this issue, median filtering and local search are used to construct outlier-free correspondences.

In the following tests, the pose+registration parameters are held constant. Since the line feature extraction is driven by the model and initial pose hypothesis, we assume a small threshold of $\tau_{mo} = 0.1$ meters, while the LADAR sample is obtained in a brute force manner with a $\tau_{mr} = 5.0$ meters. We set the omission attenuation for the CCD data at $a_{mo} = 0.5$ based upon our previous work [3]. Since we lack specific knowledge for the loss of LADAR data, we use $a_{mr} = 1.0$ for a linear omission response. Finally, for the initial tests, we assume equal importance of information coming from both sensors ($\alpha_{match} = 0.5$), since we currently lack motivation to bias the match towards one set of data.

For both of the example runs shown here, we started with the same initial pose (shown in Figures 2a and d). The initial correspondence was built based upon the initial coregistration hypothesis. In the CCD, lines with average distance less than 30 pixels and orientation difference less than 15° were included in the initial pose. For the LADAR, points within 0.5 pixels in the x and y dimensions and 10.0 meters in distance were paired. While these values provide a relatively small and mostly correct initial correspondence, an enlarged initial correspondence would include significantly more than 50% outliers and median filtering would fail.

Figures 2a and d show the initial positioning of the model features (shown in black) and the data features (shown in grey). In the final results, a similar scheme is used, with the addition that features included in the match are filled. In the optical images, the features (both model and data), included in the match are shown in grey, and the unmatched features are shown in black. The correspondence between individual features is not explicitly shown.

In Figures 2b and e, the final results of median filtering show a generally good match, indicating the absence of significant outliers. Notice that the LADAR points (Figure 2e) generated by the top of the vehicle are not included in the match, since they match poorly. Local search (results shown in Figures 2c and f) did not converge as well as median filtering. Notice in Figure 2f that not all of the poorly fitting points across the top of the vehicle have been removed. This

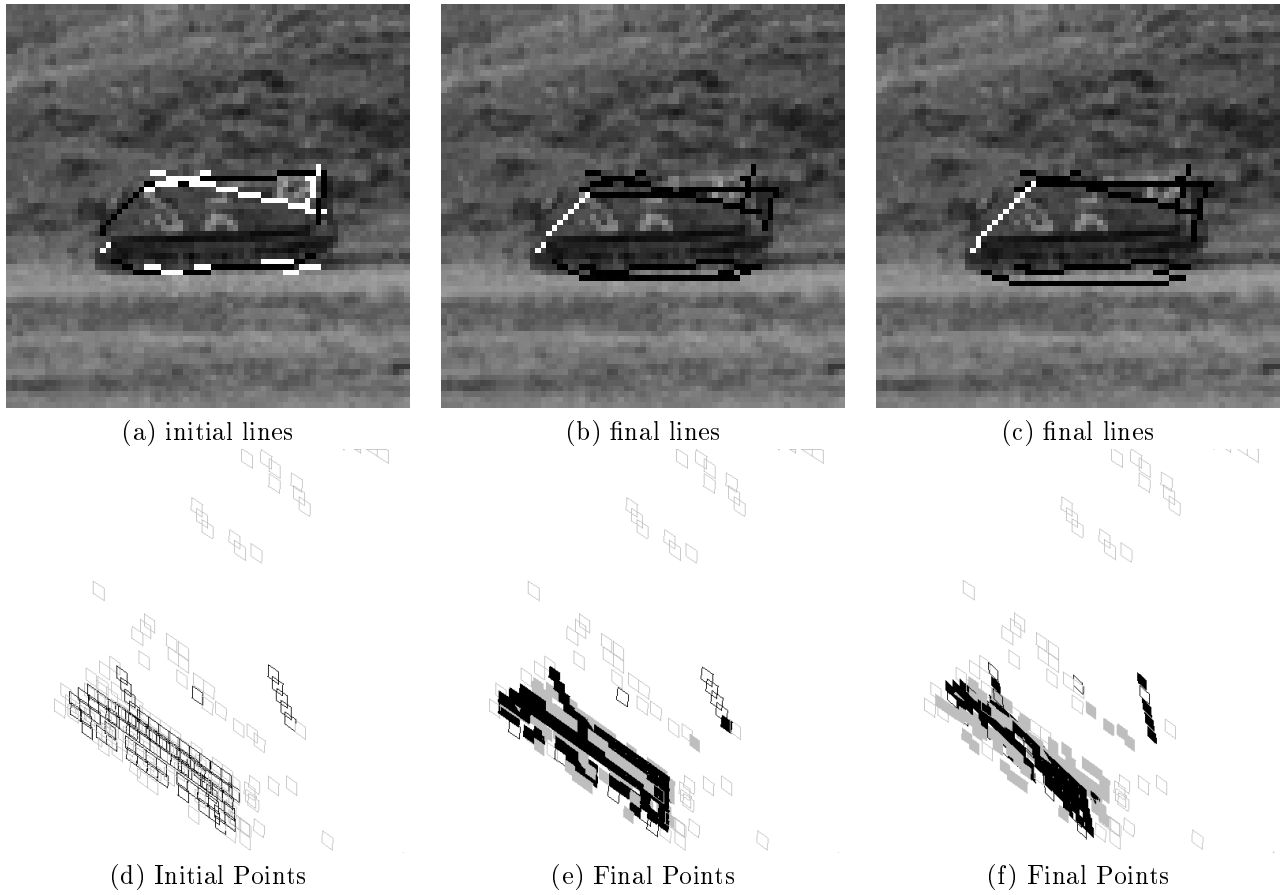


Figure 2: Results on real data. (a) and (d) are the initial estimates. (b) and (e) are the results using median filtering. (c) and (f) are the results using local search.

is due to the relative values of the fit and omission terms in the LADAR error; there was not enough fit gained to compensate for the omission of any one of these features. Such a problem could be corrected by altering the LADAR omission attenuation, in this case to denote that the small amount of data along the top of the vehicle is unreliable due to the low profile of the roof.

6 Conclusions

To help alleviate the necessity for precisely registered images, we proposed a least-squares method for computing the coupled pose+registration (coregistration) parameters for two sensors subject to coplanarity constraints. As with all least squares methods, coregistration is sensitive to outliers, so our match must be outlier-free. Both local search and median filtering are candidates for constructing these matches, and were demonstrated on real ATR data. While median filtering performed well, given modest quantities of outliers, local search became stuck in suboptimal matches due

to the search neighborhood. Future work will address grouping of range features, and many of these local minima should disappear in the local neighborhoods defined over grouped features.

7 Acknowledgements

Thanks to Jim Steinborn for proof-reading drafts of this paper and to Mark Stevens, who provided the model and image feature extraction software.

References

- [1] J. K. Aggarwal. Multisensor Fusion for Automatic Scene Interpretation. In Ramesh C. Jain and Anil K. Jain, editors, *Analysis and Interpretation of Range Images*, chapter 8. Springer-Verlag, 1990.
- [2] Anthony N. A. Schwickerath and J. Ross Beveridge. Coregistering 3D Models, Range, and Optical Imagery Using Least-Median Squares Fitting. In *Proceedings: Image Understanding Workshop*, pages 719–722, Los Altos, CA, February 1996. ARPA, Morgan Kaufman.
- [3] J. Ross Beveridge. *Local Search Algorithms for Geometric Object Recognition: Optimal Correspondence and Pose*. PhD thesis, University of Massachusetts at Amherst, May 1993.

- [4] J. Ross Beveridge and Edward M. Riseman. Optimal Geometric Model Matching Under Full 3D Perspective. *Computer Vision and Image Understanding*, 61(3):351 – 364, 1995. (short version in IEEE Second CAD-Based Vision Workshop).
- [5] 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.
- [6] J. Ross Beveridge, Rich Weiss, and Edward M. Riseman. Optimization of 2-Dimensional Model Matching. In Hatem Nasr, editor, *Selected Papers on Automatic Object Recognition (originally appeared in DARPA Image Understanding Workshop, 1989)*, SPIE Milestone Series. SPIE, Bellingham, WA, 1991.
- [7] James E. Bevington. Laser Radar ATR Algorithms: Phase III Final Report. Technical report, Alliant Techsystems, Inc., May 1992.
- [8] Richard L. Delanoy, Jacques G. Verly, and Dan E. Dudgeon. Machine Intelligent Automatic Recognition of Critical Mobile Targets in Laser Radar Imagery. *The Lincoln Laboratory Journal*, 6(1):161–186, Spring 1993.
- [9] 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.
- [10] W. Eric L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints*. MIT Press, Cambridge, MA, 1990.
- [11] 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.
- [12] Y. Hel-or. *Model Based Pose Estimation from Uncertain Data*. PhD thesis, Hebrew Univeristy in Jerusalem, 1993.
- [13] 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.
- [14] B. K. P. Horn. *Robot Vision*. The MIT Press, 1986.
- [15] Daniel P. Huttenlocher and Shimon Ullman. Recognizing Solid Objects by Alignment with an Image. *International Journal of Computer Vision*, 5(2):195 – 212, November 1990.
- [16] Alexander Akerman III, Ronald Patton, Walter H. Delashmit, and Robert Hummel. Multisensor fusion using FLIR and LADAR identification. Technical Report NRC-TR-94-052, Nichols Research Corporation, April 1994.
- [17] 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.
- [18] Rakesh Kumar. *Model Dependent Inference of 3D Information From a Sequence of 2D Images*. PhD thesis, University of Massachusetts, Amherst, February 1992.
- [19] Rakesh Kumar and Allen R. Hanson. Robust methods for estimating pose and a sensitivity analysis. *CVGIP:Image Understanding*, 11, 1994.
- [20] David G. Lowe. Fitting Parameterized Three-Dimensional Models to Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(5):441 – 450, May 1991.
- [21] 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.
- [22] Mark R. Stevens and J. Ross Beveridge. Optical Linear Feature Detection Based on Model Pose. In *Proceedings: Image Understanding Workshop*, pages 695–697, Los Altos, CA, February 1996. ARPA, Morgan Kaufman.
- [23] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*, chapter Local Search, pages 454 – 480. Prentice-Hall, Englewood Cliffs, NJ, 1982.
- [24] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.
- [25] Peter J. Rousseeuw and Annick M. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.
- [26] Anthony N. A. Schwickerath and J. Ross Beveridge. Model to Multisensor Coregistration with Eight Degrees of Freedom. In *Proceedings: Image Understanding Workshop*, pages 481 – 490, Los Altos, CA, November 1994. ARPA, Morgan Kaufmann.
- [27] 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.
- [28] Mark R. Stevens. Obtaining 3D Shilhouettes and Sampled Surfaces from Solid Models for use in Computer Vision. Master’s thesis, Colorado State Univeristy, Fort Collins, Colorado, September 1995.
- [29] J. G. Verly, D. E. Dudgeon, and R. T. Lacoss. Progress Report on the Development of the Automatic Target Recognition System for the UGV/RSTA LADAR. Technical Report 1006, Massachusetts Institute of Technology, Lincoln Laboratory, March 1995.
- [30] 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.