Computer Science
Technical Report

Coregistering 3D Models, Range, and Optical Imagery Using Least-Median Squares Fitting *

Anthony N. A. Schwickerath
J. Ross Beveridge
Colorado State University
schwicke/ross@cs.colostate.edu

December 29, 1995

Technical Report CS-96-108

Computer Science Department
Colorado State University
Fort Collins, CO 80523-1873

Phone: (970) 491-5792 Fax: (970) 491-2466
WWW: http://www.cs.colostate.edu

*This work was sponsored by the Advanced Research Projects Agency (ARPA) under grants DAAH04-93-G-422 and DAAH04-95-1-0447, monitored by the U. S. Army Research Office.

Coregistering 3D Models, Range, and Optical Imagery Using Least-Median Squares Fitting *†

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

Abstract

When using methods for recovering the least-squares optimal pose+registration parameters between a model and an image suite, construction of the feature correspondences is key. Inclusion of outliers in the correspondence set can severly deteriorate the performace and fidelity of these pose+coregistration estimation algorithms. We consider the use of median filtering in the construction of consistent correspondences. Finally, we test our coregistration with a median filtering system on real ATR data.

1 Introduction

We have been investigating sensor suite pose recovery in a multi-modal ATR domain. Our formulation considers not only the relative pose between the model and the sensor suite, but also the constrained intersensor registration. A least-mean-square-error algorithm has been developed for simultaneous estimation of both pose and registration (pose+registration) [7]. As a shorthand, the term *coregistration* is used to describe this process.

Our coregistration algorithm takes as input a set of correspondences between model features (points and lines) and an initial coregistration estimate. It then uses a non-linear optimization procedure to arrive at a coregistration estimate which minimizes a sum-of-squared-error between these corresponding features. The construction of the initial set of corresponding features is derived from initial expectations regarding the sensor registration and object pose.

As with any least-mean-square-error fitting method, the system is sensitive to outliers in the correspondence set. Local search has been shown, using features from optical imagery, to not only remove outliers, but to find optimal sets of corresponding features even when perhaps only 5% of the total candidate features match [2]. Our long term goal is an extension of this work to mulitsensor matching using the fitting procedure defined in [7]. However, our current reliance upon ungrouped points to represent range leads quickly to search spaces of intractable size. One solution is to perform some grouping on the range data, and work on this has begun. Another is to consider a more conventional approach to outlier removal: median filtering. This paper presents a least-median-squared-error extension of our previous coregistration work and demonstrates the algorithm on real ATR data.

*This work was sponsored by the Advanced Research Projects Agency (ARPA) under grants DAAH04-93-G-422 and DAAH04-95-1-0447, monitored by the U. S. Army Research Office.

†Appears also in the Proceedings of the 1996 ARPA Image Understanding Workshop.

2 Background

In our past work [7], we presented a coregistration recovery algorithm. We hypothesized that utilizing constraints between the sensors allowed for a more accurate pose estimate to be computed. We also noted that data in some multimodal domains, including ATR, tends not to be boresighted. Due to mechanical vibrations and torsions, day-to-day variations of several pixels can be expected. Our coregistration takes these inaccuracies in the sensor registration into account, applying corrections to the intersensor translation.

The least-squares algorithm developed in [7] and extended in [1] utilizes an iterative non-linear optimization method. Such algorithms require an initial parameter estimate. The fidelity of this initial pose+registration estimate partially determines whether the coregistration algorithm will converge to the correct minimum on the error surface. To investigate this problem, two sets of experiments were conducted and reported in [7]. These experiments tested the recovery of pose+registration parameters given a poor initial estimate using perfect synthetic data. It was found that the algorithm could recover, with high probability, from initial estimates up to 45° and 100 meters off.

The second experiment utilized synthetic data with Gaussian noise introduced and a perfect initial pose+registration parameter estimate. This experiment placed some bounds on the recoverability of the coregistration parameters. We found that the amount of error tolerable in the image is related to the resolution of the image. As the standard deviation of the measurement error increases, the convergence point tends to migrate. Practically speaking, this implies that the higher-resolution color data can have greater pixel-uncertainty, while the low-resolution range data needs to have relatively little pixel-error.

When we attempted to coregister real ATR images from the Fort Carson data set, we found after considerable (unpublished) exploration that our algorithm proved sensitive to both the geometry of the model and image features and also to the correctness of the correspondence set. If outliers were introduced into the correspondence set, we found convergence to neighboring minima, with large rotational errors.

Some statistical methods have been proposed in the traditional literature and have been utilized by vision researchers to construct outlier-free correspondence sets. Kumar [4] utilized median filtering to evaluate his 3-D full perspective pose recovery system and here we will adapt it to coregistration.

3 Median Filtering

Least-squares methods, such as our E_{fit} measure, assume that the data has Gaussian random noise added to it. If, however, the correspondence data contains outliers, our method will be thrown off. Median filtering is a robust statistic for detecting and removing outliers.

Median filtering [6] 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. And since the median is insensitive to up to 50% outliers, so is median filtering.

The down side is that, for non-differentiable error functions, a combinatorial search of the subset space needs to be explored. To approximate the complete combinatorial search, we can select a number of small subsets, assuming that we have a high probability of sampling at least one subset which contains no outliers. This yields the optimal fit, and allows us to throw out all data not accounted for by the Gaussian assumption (ie, 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 we would need to select at least 3 optical lines and 1 range point. However, Kumar [4] 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 stablize 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 $\texttt{cutoff} = (a \times s)^2$ where $s = \frac{\min \tilde{E}_{fit}}{0.6745}$ 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 Results and Discussion

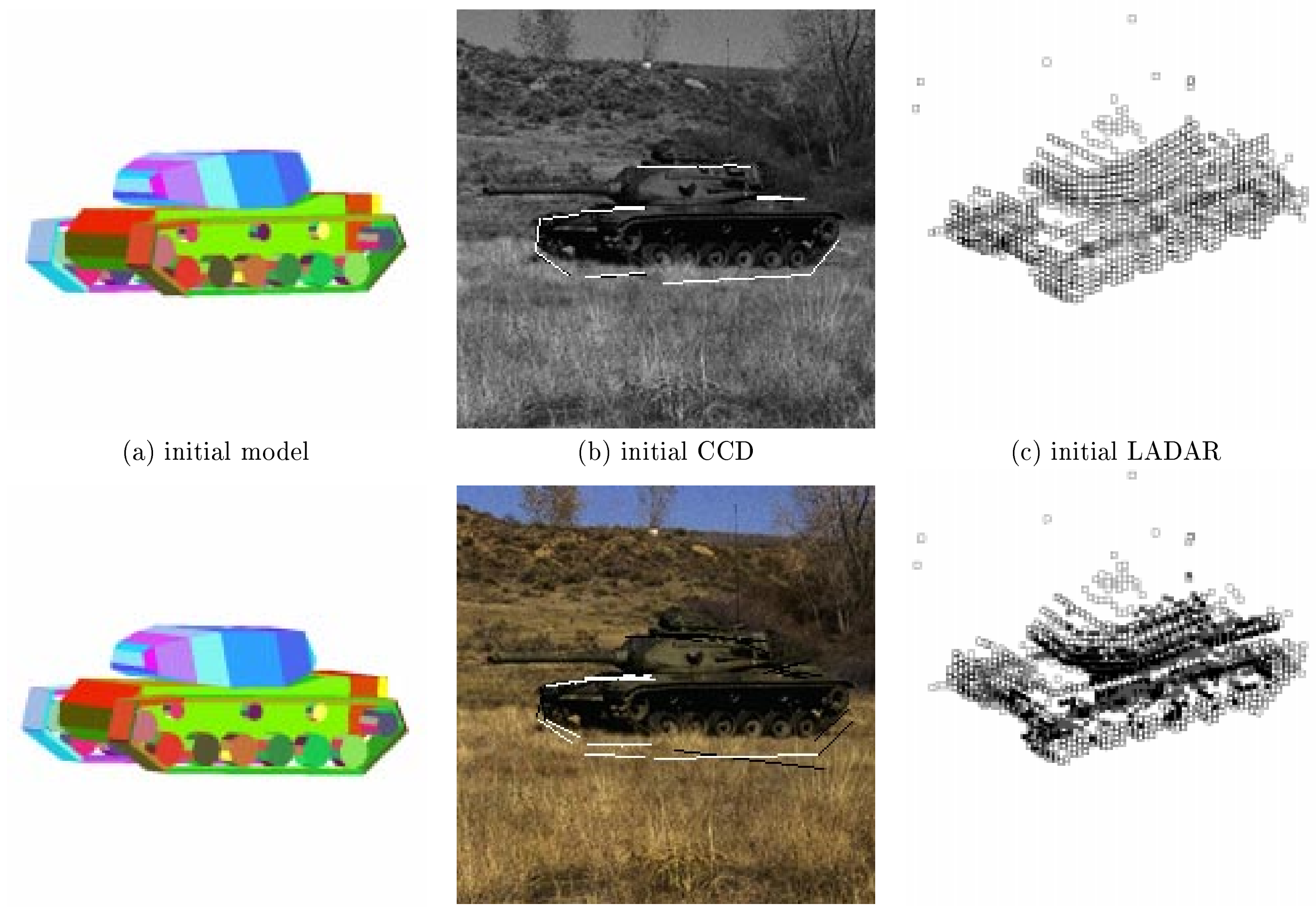
Since we wish to utilize our coregistration recovery algorithm as the alignment-and-evaluation component of a matching system, we need to have a notion of how our current system performs on real ATR data.

The initial set of corresponding model-image feature pairs, S, is selected based upon spatial proximity given a hand-picked initial pose+registration estimate. Proximity thresholds are chosen based both upon the error we observe in the feature extraction and by the percentage of outlines permitted by median filtering. We use the neighborhood $(x, y, r) = \pm(0.5, 0.5, 10.0)$ in the ladar data (x and y are in pixels and r is range in meters) and $(d, \theta) = \pm(30, 15)$ in the color data (where d is the average distance in pixels and θ is the rotational difference in degrees).

Median filtering is run on 300 subsets of 10 feature correspondences each. For random feature selections, we normalized the selection so that there is an equal probability of selecting a feature from either range or optical sensor. If this was not done, the selection would be biased towards the LADAR data (which accounts for over 95% of the correspondences $s \in S$), and the CCD portion of the error would often be ill-conditioned.

For the coregistration algorithm, we used weighting factors as described in [1] which simply replace the w_{mc} and w_{cl} of [7] with 3 intuitive factors. First is a weighting term (α) for controlling the relative importance of the sensors. Since we lack knowledge about the importance of these, we set $\alpha = 0.5$. We are also normalizing the individual sensor errors using what amounts to the second standard deviation of the presumed Gaussian noise for the sensor. We will call these terms τ_o (optical/CCD) and τ_r (range/LADAR) and set them to 0.25 and 5 meters respectively. These values seem to intuitively represent the error we are observing, though they were not found using a formal error estimation process. In order to lend additional stability, we invoke the Levenberg-Marquardt rule not only when the error would increase, but also when a rotation update of greater than 10° is proposed.

The results of the median filtering are given in Table 1. Median filtering takes on the order of 30 minutes and this time is sensitive to the total number of corresponding



(a) initial model (b) initial CCD (c) initial LADAR

(d) final model (e) final CCD (f) final LADAR

Figure 1: Median filtering results on Shot 18 (M60)

Image	CPU Time (sec)	Correspondences	
		Initial	Final
Shot 18 (M60)	1633.90	461	333
Shot 20 (M113)	2184.98	557	392

Table 1: Median filtering results

pairs $s \in S$. For each subset, pairwise error for every pair $s \in S$ must be calculated, and the pairs ranked in order of ascending error [1]. Table 1 also shows that roughly $\frac{1}{4}$ of the pairs in S are outliers: the numbers given in the last column indicate the number of pairs determined not to be outliers.

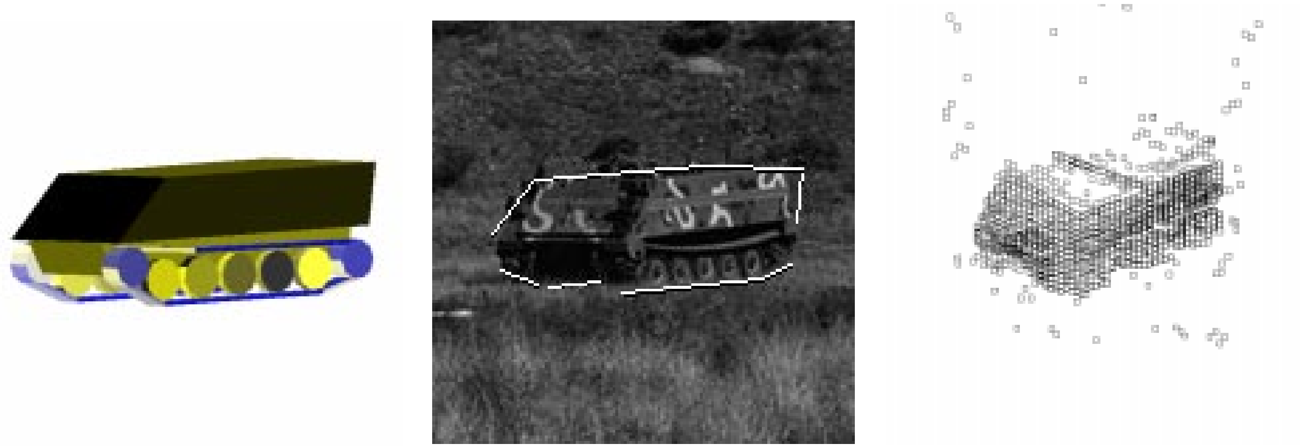
Figures 1 and 2 are initial and final coregistration estimates for two pairs of range and optical images. The leftmost column shows the target model itself. The middle column shows features for the optical imagery. The white features in Figures 1b and 2b indicate the projected 3D target silhouette. The white features in Figures 1e and 2e show both silhouette and image features determined not to be outliers: black indicates outliers. The image features are found using a model-driven approach described elsewhere in these proceedings [5]. The rightmost column shows the range data. Black squares are model range points, grey squares are LADAR pixels. Filled squares are determined not to be outliers.

The overall global change between initial and final coregistration estimates in Figures 1 and 2 is small. This is a consequence of initializing the algorithm with a good initial estimate and tight bounds upon the proximity search used to construct the set S. While the change is small, median-filtering does refine the estimates in each case, and as indicated by the numbers of outliers removed (Table 1), this fine adjustment is based upon a significant refinement of the correspondence.

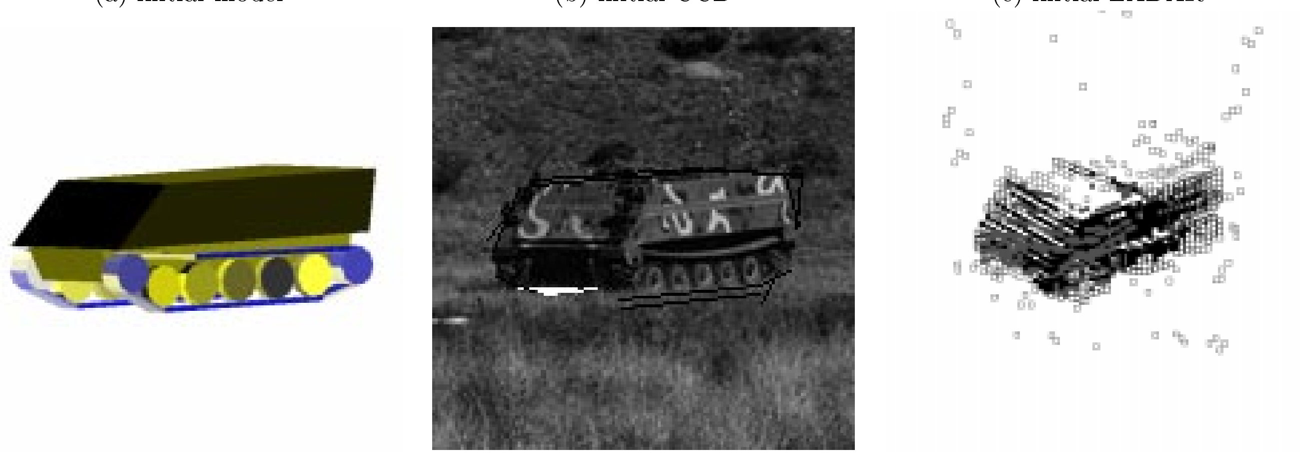
One characteristic of median filtering is that it tends to remove features on surfaces viewed from an oblique angle. Examples include the top of the M60 turret in Figure 1 or the roof of the M113 in Figure 2. A probable explanation is that such surfaces tend to have greater sampling error in the range.

In Figure 2e, note that almost all CCD feature have

[1]It should also be kept in mind that some of the computation expense relates to the C++ implementation of our algorithm. However, while a more optimized version could be implemented, the dependency on the number of features is S would still hold.



(a) initial model (b) initial CCD (c) initial LADAR



(d) final model (e) final CCD (f) final LADAR

Figure 2: Median filtering results on Shot 20 (M113)

been identified as outliers. Also note that the position of the black features suggests this is not due to an error in the feature extraction: the features are essentially in the correct positions. It is believed that for this example, the large planar surface of the M113 fits the range data so well, that relative to this fit, the CCD features are considered to be outliers.

5 Conclusion

Pose estimation methods which minimize the mean-square-error become unstable when outliers are introduced into the correspondence. We have already introduced one such method for simultaneously recovering the pose+registration parameters in [7]. The previous work, however, could not be demonstrated in a real ATR domain, due to the unavailability of automatically extracted model and data features and the inability to generate outlier-free correspondences.

As a consequence of work done in [8, 3], we are now able to extract model and data features on real ATR images. In this paper, we have constructed outlier-free correspondences using median filtering. Using the real ATR image features and median filtering coregistration, we have constructed outlier-free correspondences. Due to the expense of the current coregistration implementation, the time required to run median filtering is relatively high. However, it does mark a significant number of initially considered correspondences as outliers. We have shown that these filtered correspondences do provide stable coregistration results.

References

[1] Anthony N. A. Schwickerath and J. Ross Beveridge. Coregistration of Range and Optical Images Using Coplanarity and Orientation Constraints. In *1996 Conference on Computer Vision and Patter Recognition*, page (submitted), San Francisco, CA, June 1996.

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

[3] J. Ross Beveridge, Mark R. Stevens, and N. A. Schwickerath. Toward target verification through 3-d model-based sensor fusion. *IEEE Transactions on Image Processing*, page (Submitted), 1996.

[4] Rakesh Kumar. *Model Dependent Inference of 3D Information From a Sequence of 2D Images*. PhD thesis, University of Massachusetts, Amherst, February 1992.

[5] Mark R. Stevens and J. Ross Beveridge. 3D Model Feature Prediction for Combined Range and Optical Object Recognition. In *Proceedings: Image Understanding Workshop*, page (to appear), Los Altos, CA, February 1996. ARPA, Morgan Kaufman.

[6] Peter J. Rousseeuw and Annick M. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.

[7] 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.

[8] Mark R. Stevens. Obtaining 3D Silhouettes and Sampled Surfaces from Solid Models for use in Computer Vision. Master's thesis, Colorado State University, September 1995.