2D-3D Registration of Optical and Ladar Imagery for Real-Time Tracking

Andrew Mastin,[1,2] Jeremy Kepner,[1] John Fisher III[2]
[1]Lincoln Laboratory
[2]Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge MA 02139
mastin@csail.mit.edu kepner@ll.mit.edu fisher@csail.mit.edu

Introduction

In many target tracking situations, the sensor is moving with respect to the reference frame. A common scenario for such a setup is performing vehicle tracking with airborne surveillance. Temporary occlusion of moving objects (e.g. a car passing under a bridge) often confuses tracking algorithms, and this problem is compounded in the case of a moving image sensor. Even if a target is static, occlusion can occur due to parallax from objects between the sensor and the target. However, if 3D structure of the scene is available that can be fused with the aerial images, tracking algorithms can infer target occlusions and track targets in a more robust manner.

Recent advances in airborne laser radar (ladar) imaging technology have made the acquisition of large-scale digital elevation data more efficient and cost effective. In particular, flash ladar based on avalanche photodiodes allows for collection of numerous point returns per pulse of transmitted light. These systems are capable of collecting high resolution 3D imagery – often on the order of 30cm – as shown in Figure 1. The resulting availability of 3D scene information motivates the design of real-time systems that fuse imagery and geometry to improve tracking capabilities.

A notional example of this technology is a system involving unmanned air vehicles (UAVs) collecting video of a scene for multi-target tracking and 3D scene modeling. The UAVs would fuse the incoming video images with previously collected ladar data to infer occlusions during tracking. The most challenging computational aspect of such a system would be the real-time fusion of video imagery and ladar data. This can be decomposed into two tasks: (1) performing precise feature-based registration of the aerial imagery with the ladar data and (2) tracking multiple features that are necessary for registration in order to update the alignment with time.

Automatic Registration Methods

A variety of issues make registration of optical imagery with the ladar data a challenging task. The information offered by the two modalities is fundamentally different, as ladar collections provide 3D geometry images while photographs exhibit a 2D projection of intensity values. Effectively, traditional registration methods that use correlations of intensity values cannot be used directly. Ladar data is viewed as a projection of points onto a 2D plane, as shown in Figure 1, and points that are occluded in the optical image are visible in the ladar image. Thus, a valid registration algorithm must either infer occlusions or allow for occlusion discrepancies in a robust manner.

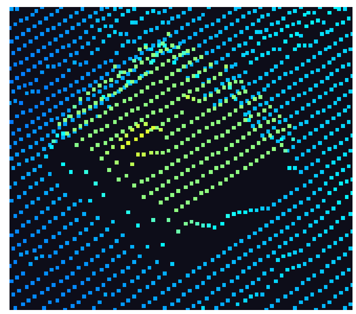

Figure 1: Ladar image of a building. Colors indicate elevation.

GPS and inertial navigation system (INS) data can often be used to find an approximate registration with respect to the ladar imagery, but near pixel accuracy is necessary for fusion of data and can only be achieved with data-driven techniques. Many previous approaches for refining registration results from GPS and INS data use correlations of features in both image modes, such as straight lines, to find an optimal registration [1,2]. They employ an exhaustive search over camera parameters (e.g. position, orientation, and focal length) and choose the minimum cost registration. This approach is extremely computationally burdensome; the system developed by Freuh, et al., for example, requires 20 hours of computing time on a conventional computer to register one image [2]. Recent approaches have led to more efficient registration algorithms, but they are not robust to all types of imagery [3].

We are currently researching a novel approach for performing registration based on information theoretic statistics and image features. Information theoretic registration methods are well suited for cases involving multi-modal imagery, as made evident by many successful developments for registration of medical imagery [3]. A variety of such methods exists, and they primarily employ joint statistical measures such as mutual information and KL-Divergence between image intensities as an objective function. Rather than using an exhaustive search, they use an optimization method to obtain a best registration. Commonly used optimization routines include both gradient optimization methods, such as steepest descent and

This work is sponsored in part by the Department of the Air Force under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and not necessarily endorsed by the United States Government.

Levenberg-Marquardt, and non-gradient algorithm such as downhill simplex and Powell's method [4].

These algorithms cannot be used directly for optical-ladar registration since the ladar data does not contain intensity values. We are instead researching the use of image features in place of intensity values to evaluate registration statistics. Our goal is to define features that (1) describe the orientation of edges, corners, and other shapes in images and (2) yield a convex objective function for camera parameters so that the optimal registration can be found reliably. Furthermore, we intend develop features that are invariant to scaling, rotation, and change in 3D viewpoint. In performing registration, this will allow us to calculate feature maps at the beginning of the registration procedure, and then find 2D projections of the ladar features corresponding to different camera poses for evaluating registration statistics. This will be more efficient than calculating ladar feature maps for each projection of interest.

Computational Aspects of Multi-Object Tracking Algorithms

In cases where sequential video images need to be registered with ladar data, it desirable to match image registration features from frame to frame so that the alignment can be updated in an efficient manner. This problem can be characterized as a conventional multi-target tracking (MTT) problem where image features are the targets and relative motion is due to the motion of the sensor.

The most preferred MTT algorithm is Multi-Hypothesis Tracking (MHT). The chief drawback of MHT is that complexity grows exponentially as the length of the tracking window grows with respect to both computational and memory requirements. Consequently, the tracker must be periodically restarted for long tracking times and/or requires the elimination of some hypotheses. Additionally, observations are not used to improve previous estimates, which is problematic in cases where two targets cross paths, or in the UAV case features cross paths due to the relative motion of the sensor.

MTT based on graphical models and belief propagation offers numerous improvements in comparison to MHT. While hypotheses are approximated via message passing, graphical model formulations, which lend themselves to distributed processing methods, do not require the elimination of hypotheses. Also, Markov structures allow for refinement of past estimates with new observations. As with MHT, complexity for exact inference is generally exponential with respect to the window length, but there are a variety of approaches for performing approximate inference that reduce complexity.

We plan to use a system similar to the one developed by Chen that uses a graphical model with parallel message-passing [6]. The model uses k-means or KD-tree clustering to maintain some knowledge of every hypothesis instead of pruning hypotheses as MHT does. Chen shows that the algorithm obtains near linear complexity with respect to the length of the tracking window. Using this algorithm in conjunction with a tractable registration procedure will provide a computationally intensive yet manageable method for performing real-time 2D-3D registration.

References

[1] A. Vasile, F. Waugh, R. Frederick, D. Greisokh, R. M. Heinrichs, “Automatic alignment of color imagery onto 3D laser range data,” *Applied Imagery and Pattern Recognition Workshop, Visualization and Transmission*, pp. 6, Oct. 2006.

[2] C. Frueh, R. Sammon, and A. Zakhor, “Automated texture mapping of 3d city models with oblique aerial imagery,” *In Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission*, pp. 396-403, Sept. 2004.

[3] M. Ding, K. Lyngbaek, and A. Zakhor, “Automatic registration of aerial imagery with untextured 3D LiDAR models,” To appear in *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2008.

[4] L. Zöllei, J. Fisher, W.M. Wells III, “A Unified Statistical and Information Theoretic Framework for Multi-modal Image Registration,” *Information Processing in Medical Imaging*, LNCS 2732, pp. 366-377, 2003.

[5] L. Zöllei, *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images*, Ph.D. thesis, MIT, 2006.

[6] M. Chen, *Efficient Multi-Target Tracking Using Graphical Models*, In preparation, M.S. thesis, MIT, 2008.