Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Real-Time Detection, Registration and Recognition using Pixel-Level Fusion of Active/Passive Imagery

Alan N. Steinberg [1] and Robert T. Pack [2]

[1] Space Dynamics Laboratory, Utah State University, North Logan, Utah, U.S.A.
asteinberg@sdl.usu.edu
[2] Civil and Environmental Engineering Dept., Utah State University, Logan, Utah, U.S.A.
rtpack@ss.usu.edu

Abstract - A system has been developed whereby active ladar and passive electro-optic imaging data are aligned in hardware at the pixel level. The resulting arrays of fully aligned, high-dimension feature data permit dense point matching in three spatial dimensions for enhanced real-time detection, registration and target recognition. Military applications for which this technology is being developed or assessed include precision tactical targeting, Precision Controlled Reference Image Base (CRIB) production and automatic registration of targeting data into the CRIB. Civil applications include 3D city modelling, real-time airborne mapping, post-disaster reconnaissance, floodplain and coastline mapping, drug interdiction target detection, environmental monitoring, and search and rescue. The system combines the ability of active systems to work at long ranges and to penetrate obscurants with the passive arrayís wide instantaneous field of view at increased resolution. This has been found to enormous benefit is in observation through partial or intermittent obscuration; e.g. with partial cloud cover or foliage.

1. Introduction

Researchers at Utah State University have developed a technique for perfectly aligning range data from a laser radar (LADAR) with a passive imaging system. The concept is illustrated in Figure 1. The key development that has enabled perfect active/passive data alignment is the electro-optical architecture that performs such alignment in the hardware.[1]

Versions of this system have been built and tested in both ground and aircraft sensor platforms. The electro-optical system architecture that accomplishes this perfect active/passive imagery alignment is amenable to a diversity of implementations, involving one or several IR or visible bands of passive imagery and selection of diverse LADAR wavelengths, modulations and optical trains.

[1] This effort was performed under internal research and development funds provided by Utah State Universityís Department of Civil and Environmental Engineering and Space Dynamics Laboratory.

Furthermore, while the initial implementations involved focal plans for the LADAR receive channel with pixel diameters an order of magnitude larger than those for the passive imaging sensor, enhanced optics can be expected to obtain high-resolution three-dimensional real-time imagery.

While the details of the system design are restricted from public release, results from the experiments have been published [1,2]. This paper extends the discussion of findings concerning the profound implications of perfectly aligned passive and active imagery data, as obtained in ongoing system experiments and in analysis.

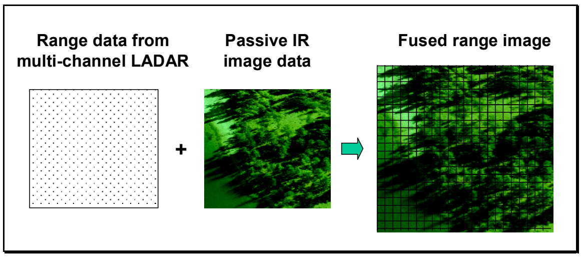


Figure 1. Pixel-level fusion of active and passive imagery data

Among benefits of perfectly aligned active and passive imagery are:

a) enhanced target detection by lowering detection thresholds in processing either or both channels of data;
b) enhanced relative and absolute registration of imagery collected by multiple sensor platforms or by a single moving platform;
c) enhanced target recognition by combining the measurements in each pixel from multiple sensing modalities;
d) using the fill three-dimensional imagery to select 2D image planes, allowing tomographic viewing of partially obscured situations.
e) The last of these benefits has been examined in previous publications[1,2]. The present paper adds discussion of the first three benefits.

2. Pre-Detection Fusion

Aligning multi-channel data at the pixel level enables enhanced detection of entities in the presence of clutter to the extent that the noise processes in the N channels are conditionally independent of one another. with full conditional independence the fused probability of detection improves as

$$P_D = 1 - \prod_{i=1}^{N} (1 - p_D(i)).$$

Therefore, detection thresholds in one or all sensor channels i may be reduced by the ratio $p_D(i)/P_D$ without sacrificing performance in terms of $p_{FA}(i)/P_D$. Thus, the high degree of signal correlation expected in pixels in which the active and passive channels are aligned will result in an enhanced signal-to-noise ratio, as the noise will tend to decorrelate between the active and passive channels. The same can be said of multi-channel active and/or passive data aligned at the pixel level.

3. Imagery Registration

The use of pixel-aligned active and passive sensors has significant implications for registering data from a moving sensor platform over time or from multiple sensor platforms.

Consider the problem posed in registering passive imagery over geometrical diversity, illustrated in Figure 2. Here points in two-dimensional images must be associated.

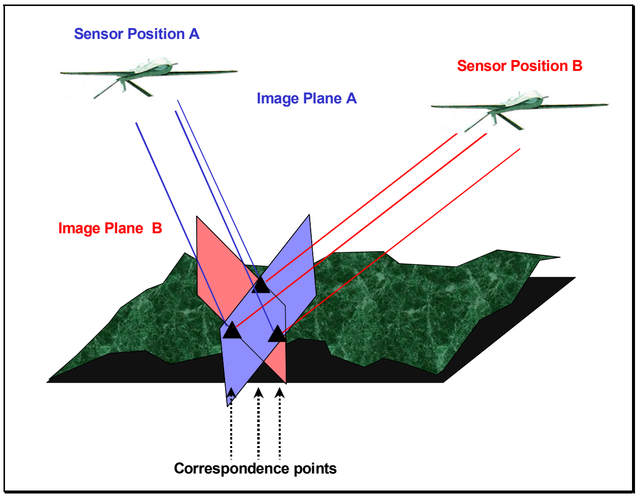


Figure 2. Registration problem in 2D imagery

Relative registration can be performed either using sparse or dense point methods. Sparse point methods involve the selection of a small number of points considered likely to be recognized across images, generally on the basis of spatial resolution, high contrast, and likelihood of being within multiple images and of persisting over the corresponding observation times. It will be noted that Absolute registration is generally implemented as a sparse point process.

Dense point methods presume no point selection, but involve correlation in two-dimensions across multiple images. Dense point processes have the obvious advantage of requiring a higher degree of correlation: typically on the order of 10^6 points, in contrast to order 10^1 for sparse point methods.

Registering in two-space images collected in three- space can incur the general problem of ìghostingî; i.e. of point misassociation across images. Figure 3 illustrates this phenomenon, in which similar features of distinct portions of the observed environment are associated.

Ghosting results from ambiguity in feature space. In sparse point association – whether for relative or absolute registration – such ambiguity is in the measurements of features in the chosen reference points. In dense point association, ambiguity is in the features of the observed scene. This can be a problem in relatively featureless terrain, in which the correlation attending to individual small entities can be overwhelmed by the low information content in the remainder of the imaged areas.

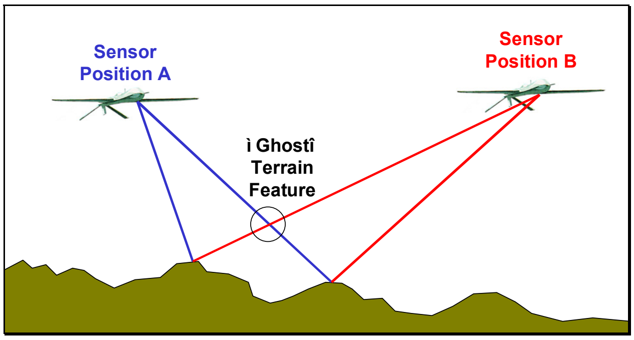


Figure 3. The problem of "ghosting" in two-dimensional data association

The pixel-level alignment of a precision LADAR with the passive imagery sensor elevates the problem to associating points in multiple images to a three-dimensional problem. In other words, the requirement become that of associating (sparse or dense) points in *intersecting projected image volumes*, rather than in *projected image planes*. Figure 4 illustrates the measurement volumes in which images acquired from two sensor positions are to be point-associated and aligned. This increases the dimensionality and, therefore, the stringency of the association requirement.

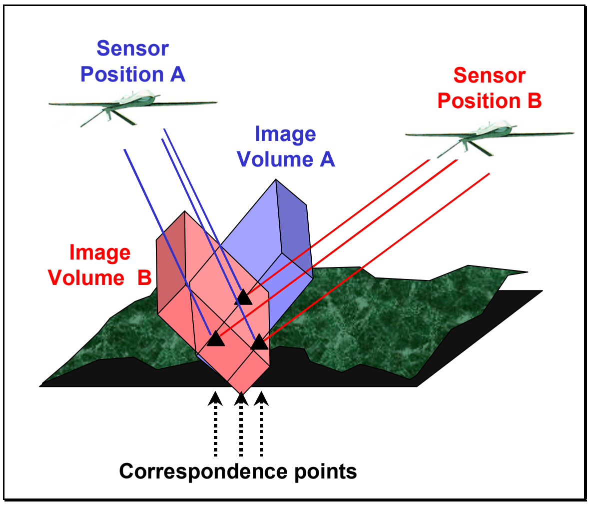


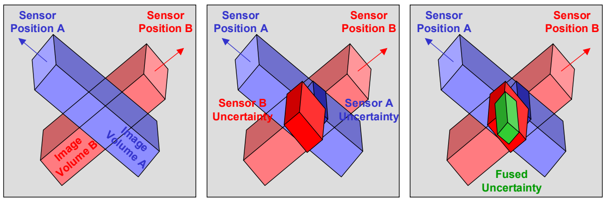
Figure 4. Registration problem in 3D imagery

Furthermore, because the LADAR measurements are most precise in the radial dimension – exactly orthogonal to the high precision passive image plane – the fully-registered 3D data from the combined active/passive sensors permits a significant improvement in 3D positional precision. When registration is performed in multiple 3D images, the geometric diversity among such images can be exploited for significantly improved point matching and, therefore, improved relative registration.

Indeed, because the LADAR used in the subject system has an effective pixel diameter that is only one order of magnitude larger than that for the passive array, dense point matching techniques can be performed in 3D. This can also improve the correlation among multiple images, resulting in improved estimation of relative and absolute biases among multiple images.

Figure 5 illustrates the concept of improved (relative) registration accuracy. Frame (a) shows the relatively precise angular and uncertain range information in the passive imagery. This allows r(relative) registration accuracy on the order of the sensor uncertainty intersection, as shown in Frame (b). In contrast, the contribution of the precise range information in each 3D image permits registration error to be reduced to the intersection of the 3D measurement, as shown in Frame (c).

This precise data association permits an enormous capability to assemble mosaics of co-registered images. Such capability is illustrated in Section 5 below, in which multiple 3-dimensional images are precisely co-registered to develop views of terrain that is partially or intermittently obscured by clouds or vegetation.



(a) Image Volumes (b) Unfused Angle/Range Accuracies (c) Fused Angle/Range Accuracy

Figure 5. Use of 3D imagery in refining positional measurement alignment

4. Target Recognition

The perfectly aligned active and passive pixel data result in an extended feature vector per pixel, incorporating the data from both sensor channels. Such extended feature pixel-ordered vectors permit target discrimination in higher dimensional feature space.

Furthermore, the intrinsic ability to register multiple three-dimensional images – whether from multiple sensor platforms or as a single platform moves – enables the recognition of targets from multiple finely-registered aspects. In this way, fused entity recognition is enhanced by the requirement that the fused data be consistent not only with a given target type, but with a target type as viewed from a set of aspect angles consistent with the known multiple viewing geometries to the target location.

5. Clutter Penetration

This section presents an approach to extracting EO data from multiple holes in clutter with the assistance of a range sensor such as LADAR that is fused at the pixel-level. The resulting 3D ìpatch dataî provides an opportunity to characterize the observability of targets in clutter and to compute the probability of target detection, given a target is hiding in the clutter. An approach to dealing with so-called negative information is also introduced along with a way to deal with the temporal evolution of target estimation and expectation given such data. For more detailed discussions of these concepts, see [1] and [2].

The probability of target detection and identification in cluttered environments is influenced by clutter density and the associated ability of a remote sensor to see through holes in clutter. When acquiring imagery from low-flying dynamic platforms such as small UAVs or ground-based sensors used by special forces, often only glimpses are possible though holes between clouds or trees or into steep canyons. For a given Electro-Optical (EO) image, only a limited number of small ground patches might be visible.

When a second image is co-registered with the first, the number and size of patches is increased. Given collection from enough viewpoints (as the sensor platform moves or as information is combined from multiple sensing platforms), it is theoretically possible to piece together enough patches to view a significant portion of the scene behind the clutter and discover otherwise obscured targets.

However, doing so with sufficient accuracy for target recognition poses a major problem unless the accurate fine details of the clutter geometry – e.g. the detailed shapes of trees and the geometry of the intervening gaps – can be determined. The compilation of such complex geometry through stereoscopic techniques is problematic when many points on the ground are only viewable through one gap seen from one viewing point only.

Sensing beneath tree canopies has been a long-standing goal of many military programs. Clutter penetrating technologies such as L-band radar suffer from relatively low spatial resolution and an inability to recognize many targets.

Ranging sensors such as LADAR cannot penetrate most clouds or trees but are capable of extracting data from gaps in such obscurants and determining the shape of objects revealed through the gaps. Where enough data are accumulated from diverse viewing points, it is possible to piece together detailed shapes of partially obscured objects and scenes. Piecing together such puzzles in real-time is a challenge and has been an area of focus of research projects such as the U.S. Defence Advanced Research Projects Agency (DARPA) Jigsaw program.

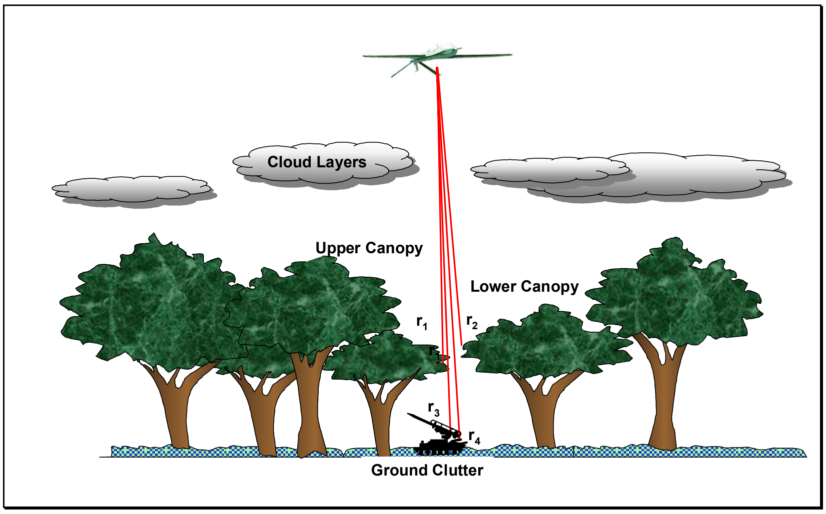


Figure 6. Concept of range data collected within gaps in clouds and tree canopy

Figure 6 illustrates this concept by showing range data being collected within a gap in multiple layers of obscuration, to include cloud layers and vegetation canopies.

EO data can be extracted from gaps in the obscuration at the same time that range data is being collected. However, because gaps open and close rapidly when the sensor is moving, it is imperative that the two data sets be collected *simultaneously*, as in the present system.

Also, when the apertures of the EO and LADAR sensors are separated on the platform, it is likely that a gap into which one sensor is seeing cannot be viewed from the other sensor. For this reason, it is important that the two data sets be collected from the *same aperture*, as in the present system.

When these two conditions are met, the confounding effects of parallax disappear so that direct pixel-level registration of LADAR to EO is possible. Figure 7 shows an example of this type of data. This figure shows two pairs of LADAR and EO imagery. Each pair was collected simultaneously and from the same nodal point (i.e. through the same aperture). On the left side of each pair is the LADAR range image, where the range values represented by individual pixels have been cyclically color-coded according their range from the sensor. For example, yellow is closer than green which is closer than cyan which is closer than blue, etc. The right-hand image for each pair is a color EO image in which 100 pixels correspond to each LADAR pixel. The pair of images on the left provides a view of a road through a hole in some brush and trees. The pair of images on the right provides a view of the same area but from a different viewing point.

A given LADAR/EO image pair can be transformed into a three-dimensional image given knowledge of the position and orientation of the sensor. Figure 8 shows a three-dimensional image derived from the fusion of the left-hand LADAR/EO image pair of Figure 7.

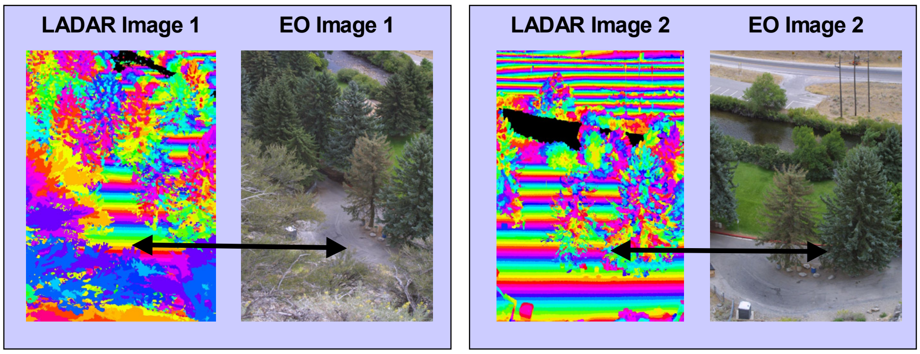
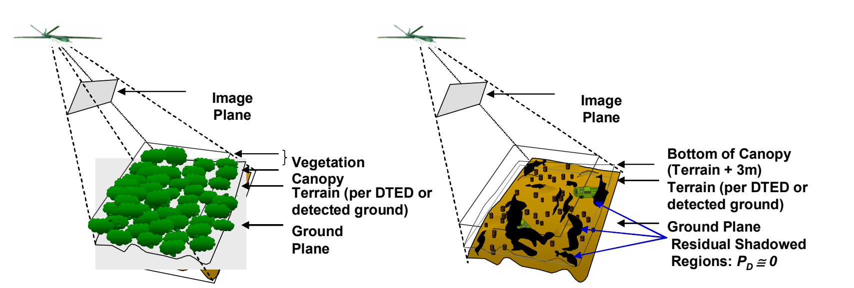


Figure 7. Two pixel-registered LADAR/EO image pairs of the same area but from different viewing points

6. Target Identification Opportunities in Clutter

The opportunity to recover range-tagged feature data through partial or intermittent obscuration permits partial 3D scene reconstruction that can be successively augmented over time. The use of multiple viewpoints enables the filling in of the inevitable residual shadowing associated with a given viewpoint. One potential visualization technique is that of selective layering. Using this technique, the analyst is able to scan or step through layers of the data, either in the range dimension or in a dimension contoured to modelled terrain. In this way, for example, the analyst can explore the scene at or near ground level by suppressing the intervening cloud or foliage obscuration. Alternatively, he can isolate aircraft or other objects from an earth background.

Figure 8 illustrates the concept of stepping through layered data. Frame 8(a) illustrates a cluttered scene. Frame 8(b) illustrates that, given knowledge of the ground surface (e.g. by means of a DTED model), the clutter can in effect be ìsliced offî at a given distance above the ground plane to expose objects near the ground. As illustrated, portions of the ground can be exposed, but other portions will remain obscured. These obscured areas are referred to as residual shadowed regions where the probability of target detection is effectively zero.



(a) Viewing partially obscured scene (b) Removal of obscuring layers

Figure 8. Slicing away layered data to reveal objects near the ground surface

This technique is demonstrated in Figure 9. Frame 9(a) shows an oblique passive electro-optical image (the same as in the left side of Figure 7). The pixel-aligned LADAR data has been used to create a three-dimensional image. That 3D image was then ortho-rectified and data from the tree canopy removed, revealing the unobscured patches near ground level, as shown in Frame 9(b). The residual shadowing is inevitable in a 3D image taken from a single viewpoint.

The technique of using multiple 3D images to expand the view through obscuration is demonstrated in Figures 10 and 11. In Figure 10, the pair of 3D images used in Figure 7 have been ortho-rectified, densely registered and sliced at 3 meters above the local ground level. Figure 11 is an expanded view of the lower left corner of Figure 10. Frame 11(a) is a birds-eye view of the three-dimensional image without any layer slicing. Frame 11(b) shows the same scene, but with ìremovalî of all clutter higher than three meters above the ground. Note how two small barrels (our targets in this case) are revealed from under a tree. These two barrels are obscured in 11(a) but revealed in 11(b). Ultimately this slicing concept can be expanded to provide full tomographic capability in any image plane.

Such capability has applications in dense, complex three-dimensional scenes as might be encountered in urban environments. E.g., it is possible to construct detailed imagery or video of individual vehicles or people as they move through crowded streets, which intermittent blockage by other vehicles or people, buildings, etc.

(a) Oblique EO Scene (b) Orthoprojected Using LADAR ñ Trees Removed

Figure 9. Example of canopy slicing

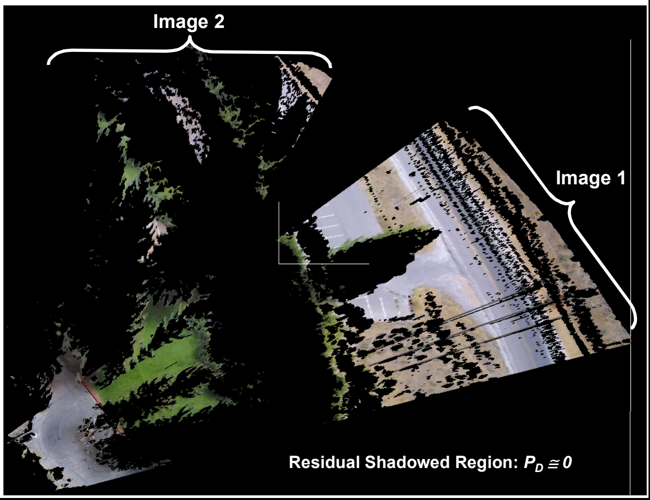
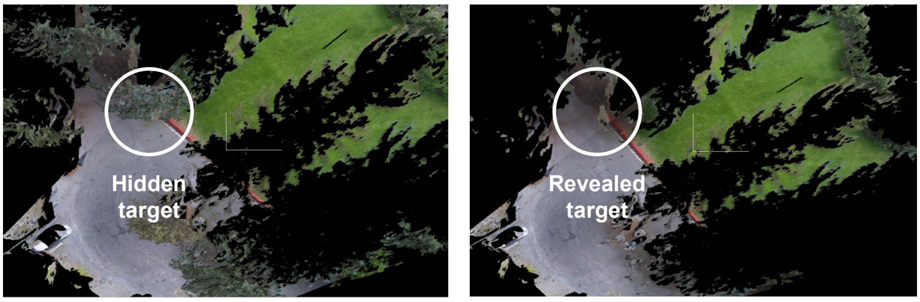


Figure 10. Combining two ortho-projected 3D images



(a) Ortho-projected image with object obscured by trees

(b) Trees ìremovedî to reveal object

Figure 11. Example of the use of tomographic slicing in a three-dimensional imagery to expose targets near the ground surface

7. Conclusions

A very high degree of compensation for motion in the sensor platform is achievable when precise range measurements are obtained from a LADAR sensor synchronized with fully co-registered precise angle data from a passive EO sensor. This precise spatio-temporal alignment of multi-sensor data eliminates the parallax problems encountered by passive systems alone and enables the real-time collection of three-dimensional imagery. Exploitation of such imagery allows target tracking and identification in clutter and the characterization of the target detection probabilities given the clutter. The probabilities of detection can be used to better allocate resources to reduce overall resource cost and increase the utility of predictive sensor management.

References

1. Pack, Robert T. and Steinberg, Alan N., ìPixel-level fusion of active/passive data for real-time composite feature extraction and visualizationî, *Proc. Workshop on Massive Military Data Fusion and Visualisation*, NATO Research and Technology Agency, Information Systems Technology (IST) Panel, IST-036-RWS-005, Halden, Norway, October 2002.
2. Steinberg, Alan N. and Pack, Robert T., ìPixel-level fusion for real-time composite feature extraction and visualizationî, *Proc. Sixth International Conference on Information Fusion*, Cairns, Queensland, Australia, July, 2003.