Persistent Electro-Optical/Infrared Wide-Area Sensor Exploitation

Andrew P. Brown*, Michael J. Sheffler, Katherine E. Dunn

Toyon Research Corp., 6800 Cortona Drive, Goleta, CA, 93117

ABSTRACT

In this paper, we discuss algorithmic approaches for exploiting wide-area persistent EO/IR motion imagery for multi-sensor geo-registration and automated information extraction, including moving target detection. We first present enabling capabilities, including sensor auto-calibration and automated high-resolution 3D reconstruction using passive 2D motion imagery. We then present algorithmic approaches for 3D-based geo-registration, and demonstrate and quantify performance achieved using public release data from AFRL's Columbus Large Image Format (CLIF) 2006 data collection and the Ohio Geographically Referenced Information Program (OGRIP). Finally, we discuss algorithmic approaches for 3D-based moving target detection with near-optimal parallax mitigation, and demonstrate automated detection of dismount and vehicle targets in coarse-resolution CLIF 2006 imagery.

Keywords: motion imagery, video, geo-registration, multi-sensor registration, 3D reconstruction, automated moving target detection, parallax, clutter suppression

1. INTRODUCTION

A large number of EO/IR intelligence, surveillance, and reconnaissance (ISR) sensors are being deployed, on platforms including Wasp through Predator and Hummingbird-class unmanned aerial vehicles (UAVs), a variety of manned platforms, and the relatively new class of Wide Area Persistent Surveillance (WAPS)/Wide Area Motion Imaging (WAMI) platforms such as Angel Fire/Blue Devil, Gorgon Stare, and ARGUS-IS/IR. In a Layered Sensing framework, a subset of these sensors may be used to simultaneously observe a region of interest to provide complimentary capabilities, including multi-band measurements, perspective diversity, and/or improved resolution for improved target discrimination, identification, and tracking. However, to optimally exploit the complimentary capabilities of these sensors, fine-grained registration of the multi-sensor data is required. Furthermore, *geo*-registration of the multi-sensor data is desired for providing accurate context to human operators, high-confidence correlation with other intelligence sources in a GIS framework, and accurate information for weapon targeting.

WAMI sensors can be leveraged to enable successful multi-sensor registration, based on the large area covered by WAMI sensors. For example, data from narrow field-of-view (FOV) sensors with non-overlapping FOVs can be registered to the WAMI sensor data, providing registration of the multi-sensor narrow FOV data in a relative coordinate system. If the WAMI data is accurately geo-registered, then geo-registration of multi-sensor data is enabled by registering with the WAMI data, which is collected concurrently with other ISR data (reducing challenges due to image intensity variations). Furthermore, due to the *persistent* nature of WAMI sensing (providing surveillance of a city-sized region for days, or longer), it is possible to build up knowledge of the scene structure to aid in geo-registration and automated information extraction processes, as discussed in this paper.

In the current state-of-practice, geo-registration of ISR imagery is typically performed based on GPS/INS measurements of sensor position and orientation (as well as sensor internal parameters), combined with a coarse-resolution DTED terrain model; however, geo-registration errors of many tens of meters may result due to the effects of relatively small orientation errors combined with large sensor-to-scene ranges. Significantly better location accuracy is desired for targeting applications, as well as for reducing biases in tracks and other information extracted from multiple sensors, enabling effective fusion. Therefore, it is necessary to rely upon the EO/IR motion imagery data itself to derive improved sensor location and orientation parameters; for example, via comparison with orthoimages from satellite-based collections, where accurate geo-registration has been provided via human-in-loop supervision. However, the new sensor data and the reference images are rarely ever taken from the same perspective, leading to difficulties due to occlusion and parallax (including non-linear and discontinuous variations in ground sample distance across images, due to 3D discontinuities such as building/terrain edges). Analogous effects occur in formation of 2D radar images. To address these challenges, we have developed geo-registration algorithms designed to optimally mitigate these effects based on 3D models of the scene. In this paper, we discuss algorithmic approaches for 3D-based geo-registration, and demonstrate

*abrown@toyon.com; phone: 805-968-6787; fax: 805-685-8089; toyon.com

and quantify performance achieved using public release data from AFRL's Columbus Large Image Format (CLIF) 2006 data collection and the Ohio Geographically Referenced Information Program (OGRIP).

In addition to geo-registration, there are many important Air Force applications for 3D scene modeling. The availability of high-resolution 3D models of a scene would enable numerous advantages for ISR applications, including mission planning, battle damage assessment, accurate line-of-sight checks in sensor resource management, and new visualization options for human operators and commanders. However, sources of 3D information such as DTED and LIDAR collections which have been geo-registered via calibration and human-in-loop procedures are not available in sufficient resolution for most parts of the globe, and may not provide up-to-date representations of human cultural features such as buildings, roads, bridges, etc. With the planned proliferation of WAMI EO/IR sensors, as well as large numbers of UAV-based EO/IR sensors with narrower fields of view, an attractive possibility is to automatically generate high-resolution, frequently-updated 3D models of large regions of interest using the available 2D EO/IR images.

Another important application of automated 3D modeling, which we consider in this paper, is moving target detection. In urban scenes containing tall buildings, and to a lesser—but significant extent—in rural scenes containing trees, hills, and other tall objects, *parallax* effects pose the greatest challenge for reliable (high detection probability, low false alarm rate) detection of moving vehicles and dismounts. In this context, parallax refers to the phenomenon by which stationary clutter objects move through the EO or IR image plane at different rates depending on the distances of the objects from the moving sensor platform. Although a number of 2D image processing techniques have been developed to partially mitigate parallax effects in moving target detection, errors due to parallax can only be optimally removed by using a 3D model to predict the motion of each stationary object from one image to the next, enabling successful separation of moving foreground targets from the stationary background clutter. In this paper, we discuss algorithmic approaches for 3D-based moving target detection, and demonstrate automated detection of dismount and vehicle targets in coarse-resolution CLIF 2006 imagery.

The remainder of this paper is organized as follows. In Section 2, we discuss automated 3D reconstruction from passive 2D EO/IR imagery, along with automated calibration of the EO/IR sensors. In Section 3, we present 3D-based geo-registration algorithms and provide example results and error analysis. In Section 4, we discuss 3D-based moving target detection with near-optimal parallax mitigation, and provide example results. In Section 5, we provide a discussion of conclusions, and acknowledgements and references follow.

2. AUTOMATED 3D RECONSTRUCTION

Automated 3D reconstruction from passive 2D EO/IR images is a problem that has been extensively investigated, but that for which a satisfactory solution has been elusive. In particular, many approaches are not tractable when applied to real-world problems (due to computational expense). And for other approaches, accurate range estimation/surface reconstruction in low-texture regions, such as on the sides of buildings, rooftops, streets, fields, etc., has not been reliably achieved due to ambiguities that result when searching for multiple-view point correspondences. Toyon has developed a novel solution to this problem, with the following properties:

- Accurate modeling of initial and intermediate reconstruction ambiguities due to occlusion effects and low-texture/homogeneous-intensity regions (the algorithm is able to accurately reconstruct these surfaces after processing multiple frames of data). The algorithm is based on particle filtering[12] in a Bayesian estimation framework to represent intermediate and final 3D reconstruction errors, including non-Gaussian distributions. This enables self-understanding of algorithm accuracy, and calculation of any desired error metric, e.g., 90% cylindrical error or full covariance matrices for each surface point.
- Completely *dense* 3D point reconstruction at the *pixel level* (more than 100 times more densely than algorithms which operate at the feature [e.g., point or line] level, enabling significantly improved surface modeling accuracy).
- In contrast to algorithms that perform 3D modeling in a voxel space, the algorithm that we have developed provides a significant improvement in computation and storage efficiency, since surfaces are automatically identified and processing is not wasted on empty space or unobservable space inside buildings or under ground.
- Well-suited to cost-effective and efficient massively-parallel implementation on commercial off-the-shelf (COTS) central processing units (CPUs) and graphics processing units (GPUs), effectively leveraging the large investment by the computer gaming consumer market to develop this affordable, ubiquitous hardware-accelerated computer architecture. Our current massively-parallel implementation is able to process greater than

1 megapixel / sec. using a mid-range PC equipped with an Intel Core i7-2600 CPU @ 3.40GHz, NVIDIA GeForce GTX 580 GPU, and 12.0 GB RAM @ 1333 MHz.

- *Generation* of terrain models in a variety of standard formats, including DEM and DTED.

Figure 1 provides an example 3D model result, which was reconstructed completely automatically by the Toyon-developed algorithm, using one complete orbit (900 frames) from camera 2 in the CLIF 2006 public release data. In this region of the CLIF scene, the imagery ground sample distance was approx. 0.4 m. In the figure, we display the 3D model in the form of a triangular mesh 3D surface model, where the color map indicates altitude, with the progression being from blue to red for lower to higher altitude. The mesh was formed with vertices at approx. 1-m spacings, and was estimated using dense irregularly-spaced 3D point estimates reconstructed using the Toyon-developed 3D reconstruction algorithm.

2.1 EO/IR Sensor Auto-Calibration

To perform automated 3D modeling using 2D video sensor measurements requires highly accurate estimation of the sensor internal (focal length, pixel spacing, radial lens distortion, etc.) and external (6-DOF position and orientation) parameters necessary to specify the projective sensor model[3]. If the internal parameters are known, the camera is said to be *calibrated*, and it is common practice to use pre-mission measurements to calibrate cameras. The calibration can be further refined via automated video processing during the mission, as included in our solution. This is beneficial, since sensor internal parameters such as focal length can be modified by temperature changes[4]. It is furthermore necessary to have accurate knowledge of the sensor translation and rotation, relative to some coordinate system, e.g., geodetic coordinates measured by GPS. Measurements of frame-to-frame variations in these 6 external parameters are assumed to be available from onboard IMU measurements, which are fairly accurate from frame to frame, but contain significant drift over time. Because of this drift, video data-driven optimization of the external parameter estimates is required, and is performed via Structure from Motion (SfM)[5], by jointly estimating the relative 3D positions of naturally-occurring scene feature points and the camera position and orientation.

To avoid the otherwise likely event that small errors in SfM-based sensor 6-DOF position and orientation errors will accumulate over time to unacceptable levels, we have developed and implemented loop-closing auto-calibration processing which operates on keyframes (e.g., 0.1 Hz) and is designed to ensure that the auto-calibration solution is consistent around an entire orbit. Between keyframes, recursive SfM processing of consecutive frames collected at the full frame rate is used to interpolate between the keyframes, providing a good compromise between auto-calibration accuracy and run time. The loop-closing auto-calibration processing is based on our development of perspective-corrected image feature extraction and matching followed by complete-orbit sparse bundle adjustment. The optimization (including perspective-corrected feature extraction, matching, and sparse bundle adjustment) is performed in multiple iterations (2-3) to bootstrap from limited initial information (sensor metadata, a planar terrain model, and images) to more complete and accurate information (e.g., optimized sensor models, automatically computed terrain model, and images). In each iteration, the accuracy of perspective-corrected feature extraction and matching is improved, which leads to improved bundle adjustment results. We have found the resulting camera position and orientation estimates to be internally consistent around an entire orbit such that *relative* registration errors are on the order of a single pixel ground sample distance (GSD). This is an important enabling capability for 3D reconstruction using a complete orbit of data, as shown in Figure 1. We have also shown[6] that this technique can be extended to perform bias estimation, including estimation of fixed sensor-to-airframe orientation parameters.

3. 3D GEO-REGISTRATION

Possible sources of reference data for geo-registration include 2D satellite ortho images and 3D LIDAR point clouds. In both cases, accurate geo-registration of the reference data was likely based on careful calibration procedures combined with human-in-loop supervision and intervention to establish accurate control point correspondences. The cost of the reference generation process limits the frequency with which reference data is updated. Challenges for geo-registration of EO/IR imagery using reference data thus include differences in appearance between the current mission imagery and reference imagery, which may have been collected using sensors with significantly different characteristics, at different times of day, and/or during different seasons. Scene content, including buildings, roads, and vegetation may have changed due to human activity and/or natural processes. Also, it is important to note that the new sensor data and the reference images are rarely ever taken from the same perspective, leading to difficulties due to occlusion and parallax. The geo-registration algorithmic approaches discussed in this section are designed to address these challenges.

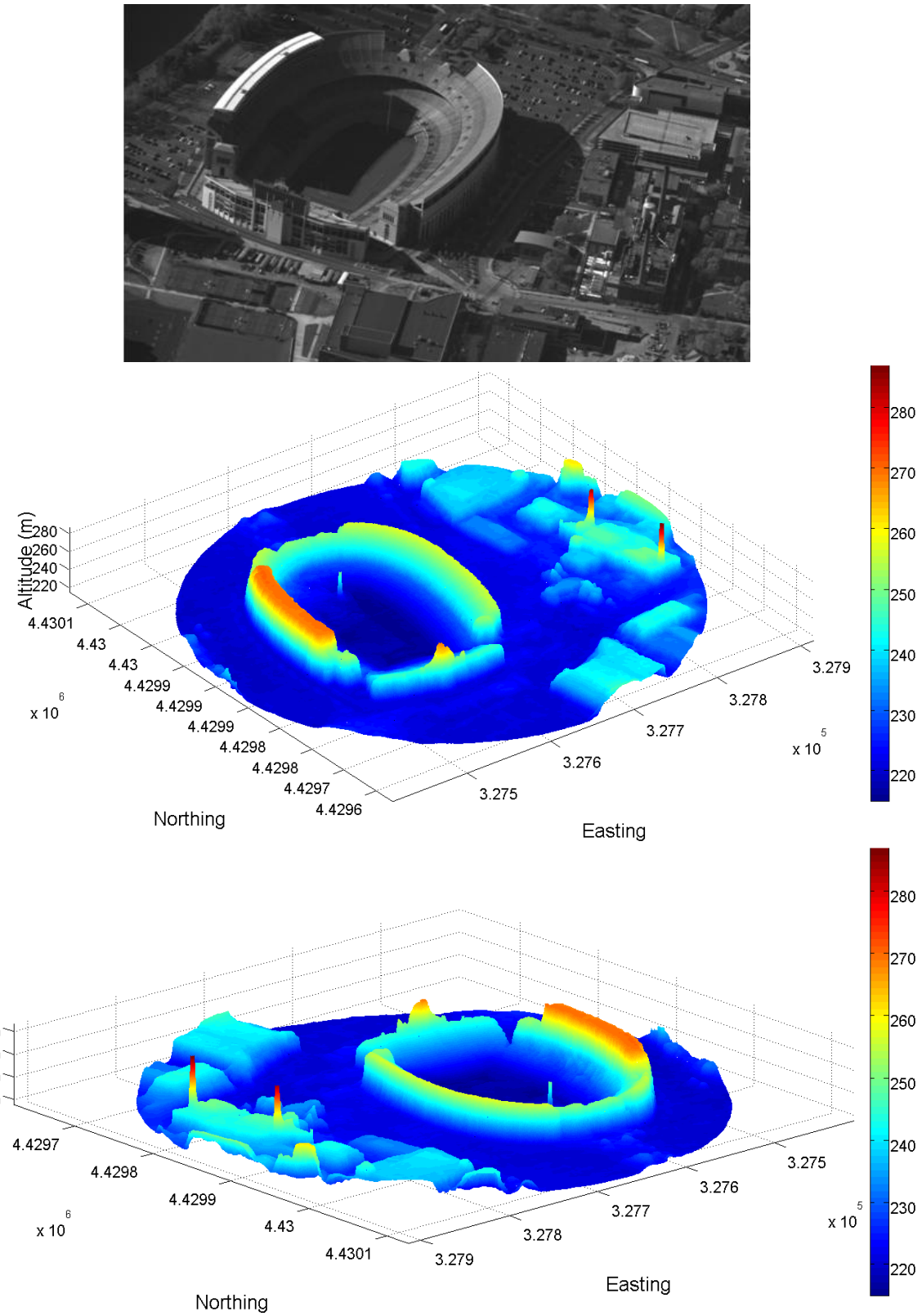


Figure 1. Example automated 3D reconstruction from passive 2D motion imagery. The upper left of CLIF 2006, camera 2, frame 1300 is shown at top, and two views of the reconstructed triangular mesh 3D surface model are shown below.

When only 2D reference data (e.g., satellite ortho-image) is available, the best we can do for geo-registration is to use the 3D terrain model that we've constructed to perspective-correct the current mission imagery to the same perspective (e.g., orthographic) as the reference image. We refer to this as a 3D-to-2D geo-registration method. When using multiple mission images collected from different perspectives, missing data (due to occlusion) can be filled in, generating a synthetic image that best agrees with the reference image. We do this by estimating a texture distribution on the triangular mesh 3D surface model. The textured 3D surface model can then be rendered to any perspective. In Figure 2, we demonstrate rendering to two example perspectives, including orthographic, which is the most common perspective for enabling geo-registration via alignment with a 2D reference image. Following perspective correction, we then perform 2D-to-2D image registration optimization using robust statistical techniques designed to optimize statistical correlations between disparate datasets.

When external reference 3D data is available, e.g., from LIDAR, direct 3D-to-3D alignment can be performed, as illustrated in Figure 3. Direct 3D-to-3D processing removes challenges due to intensity variations between the mission data and reference data. However, variations between datasets do exist and must be addressed algorithmically. One source of mission-reference data discrepancies is data resolution. For example, the OGRIP LIDAR data shown in Figure 3 is available at 7-ft post spacing in Easting and Northing, while the Toyon-generated 3D model is available at higher resolution, similar to the EO/IR sensor GSD, which for CLIF 2006 is approx. 0.25 – 0.5 m, and was approx. 0.3 m for the region shown in Figure 3. Another significant variation between the automatically reconstructed 3D model and the LIDAR-measured 3D model is that the reconstructed model includes points on sides of objects (building walls, for example), while the LIDAR model only contains measurements of the tops of objects. Finally, another source of variation is reconstruction errors and scene structure variations (e.g., a construction project was underway while the CLIF 2006 collection was performed).

To address these challenges, while maintaining computation efficiency, we have developed the following algorithmic approach, which accounts for resolution variations and includes robust statistical processing to account for 3D structure variations. The approach includes a coarse-to-fine optimization strategy to provide fine-grained registration while limiting runtime. The approach also successfully de-couples estimation of altitude registration from Easting/Northing registration, providing more than an order of magnitude improvement in computational efficiency. The processing flow includes the following steps:

- For both 3D point cloud data sets (automatic-2D-to-3D-reconstructed and LIDAR) that we desire to register:
  - Prepare an Easting-Northing 2D UTM grid with a defined spacing, `DeltaUtm` (e.g., 1 m).
    - Create two images (matrices), `Altitude` and `AltitudeStd`, corresponding to this grid.
    - Create a binary image mask (matrix), `Mask`, corresponding to this grid.
  - For each UTM grid element, find those points within some distance threshold, `DistThresh` (e.g., 3 m; selected based on LIDAR point spacing).
    - Calculate the weighted mean altitude, with weighting of points based on Easting-Northing distances from the center of the grid element. Store the weighted mean altitude in the appropriate `Altitude` element. Example `Altitude` images are shown in Figure 4(a)-(b).
    - Calculate and store, in `AltitudeStd`, the weighted mean altitude standard deviation, with weighting of points based on Easting-Northing distances from the center of the grid element. Example `AltitudeStd` images are shown in Figure 4(c)-(d).
    - If no points fall within the distance threshold, set the `Mask` element to 0, indicating that the element should not be considered in registration processing; otherwise, set the `Mask` element to 1.
- Calculate a combined `Mask` as a binary `AND` of the reconstructed and LIDAR data masks.
- Register the reconstructed and LIDAR `Altitude` and `AltitudeStd` images such that *mutual information* is maximized[7]. Mutual information describes the degree of statistical correlation between two vectors, enabling successful alignment even in cases of amplitude reversals and one-to-many amplitude mappings between the two vectors. In this case, it also provides invariance to altitude shifts between the two datasests, allowing us to decouple altitude optimization from Easting-Northing optimization. The mutual information metric is the sum of the mutual information values computed between reconstructed and LIDAR altitude images and between reconstructed and LIDAR altitude standard deviation images. The registration optimization is performed over 2D translation, followed by local optimization over scale, rotation and translation. Finally, optimization over altitude is performed based on a mean altitude difference magnitude metric.

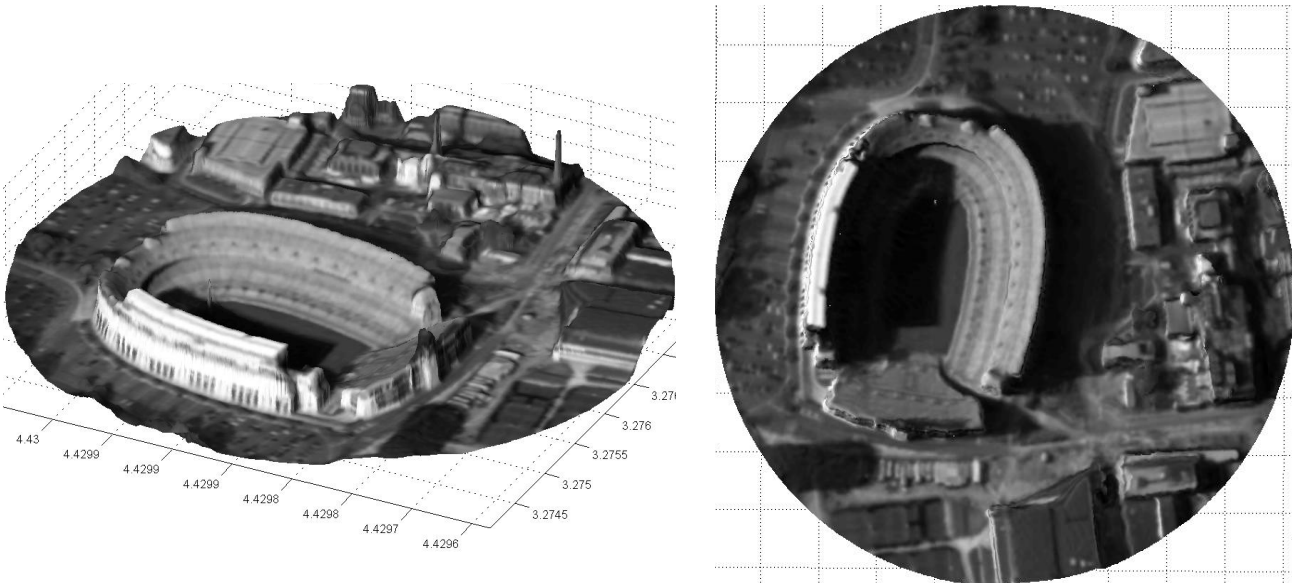


Figure 2. Example texture model for the surface of an automatically reconstructed triangular mesh 3D model. The textured 3D surface model has been rendered to two different 3D perspectives, including orthographic (right), demonstrating the ability to perspective-correct imagery prior to geo-registration processing using 2D reference imagery.

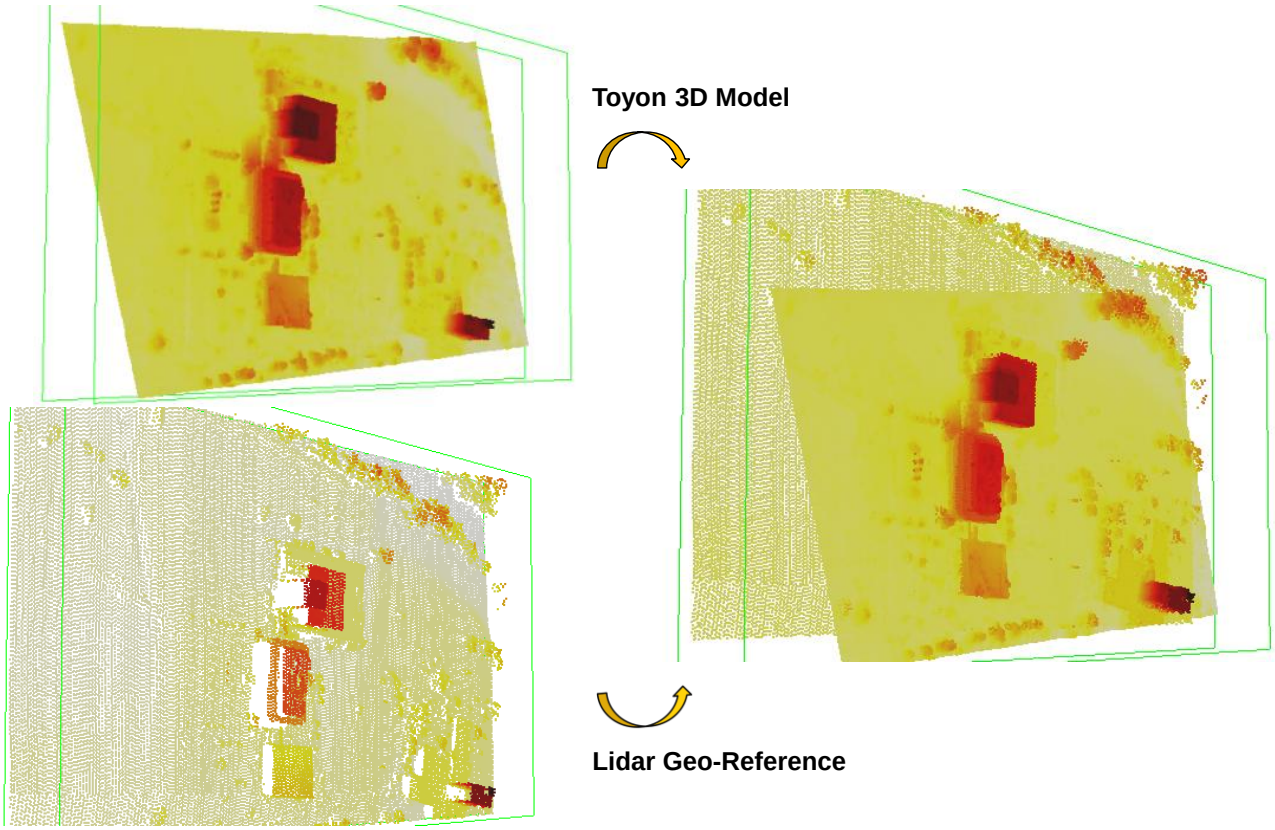
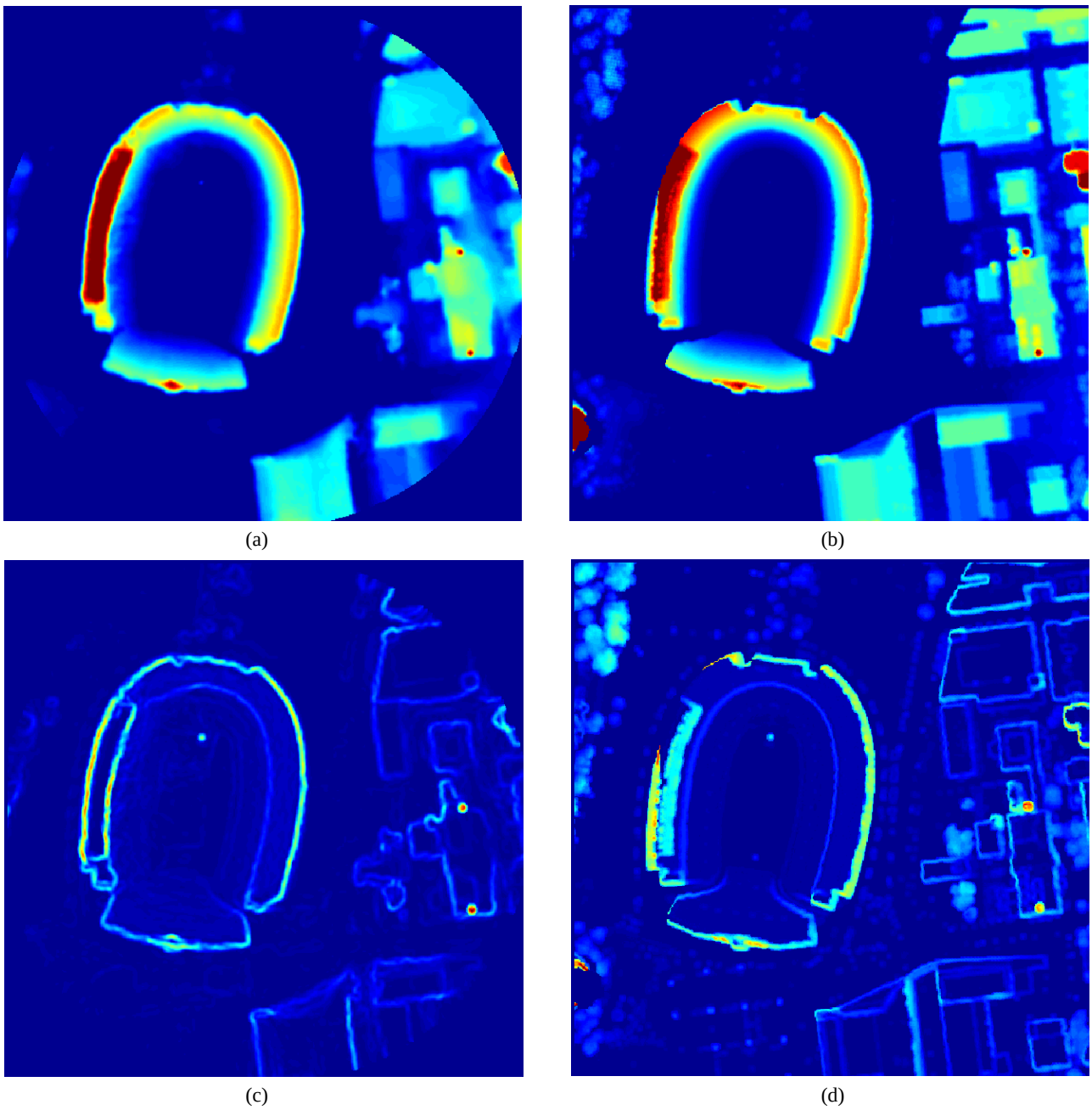


Figure 3. Example point clouds reconstructed from passive 2D imagery (top) and measured by LIDAR (bottom), and illustration of point cloud alignment following 3D-to-3D registration processing (right).



(a) (b)

(c) (d)

Figure 4. Interpolated altitude images (top) and altitude standard deviation images (bottom) computed from a reconstructed point cloud (left) and a LIDAR-measured point cloud (right) collected at different point resolutions.

To quantify the accuracy of the reconstructed geo-registered 3D model, we computed the statistical distribution of minimum distances from each of the 55,000 LIDAR points in the region to the reconstructed triangular mesh 3D surface model. Computation of the minimum distance for each LIDAR point was performed by locating the closest three points (vertices) in the reconstructed triangular mesh, and computing the minimum distance to the triangular facet defined by these vertices. The resulting minimum distance distribution is plotted as a histogram in Figure 5. The rare, larger errors in the distribution are likely due to reconstruction errors and scene differences between the CLIF and OGRIP data collections. The mean of the minimum distance distribution was 1.26 m, and the median was 0.74 m. This error is small compared with the 7-ft postings between LIDAR points, on which geo-registration of the reconstructed 3D model was based.

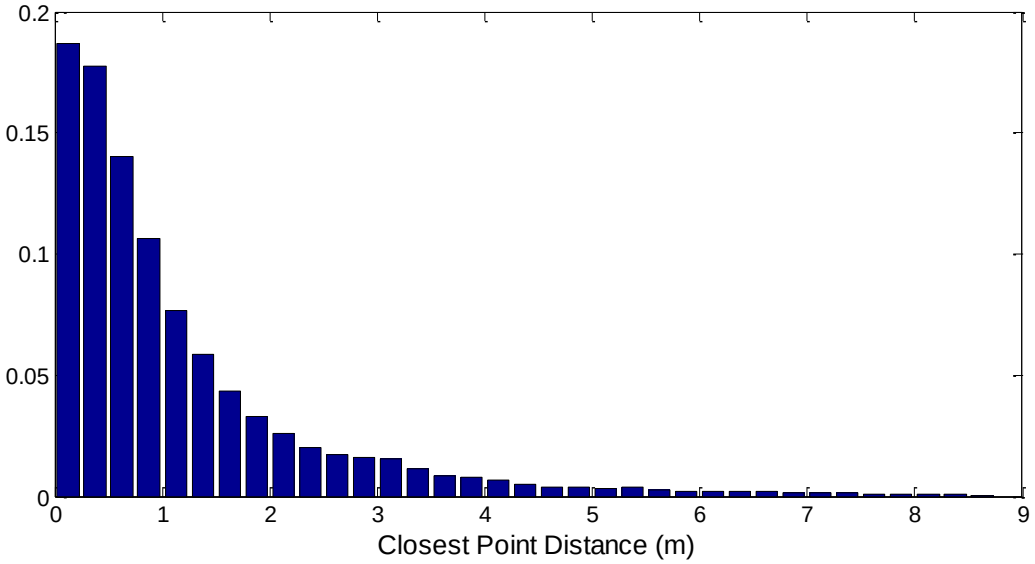


Figure 5. Histogram of minimum distances between 55,000 LIDAR points and the automatically reconstructed triangular mesh 3D surface model. The mean and median of this distribution are 1.26 m and 0.74 m, respectively.

4. MOVING TARGET DETECTION

One important application of automated 3D modeling is moving target detection and tracking. In urban scenes containing tall buildings, and to a lesser—but significant extent—in rural scenes containing trees, hills, and other tall objects, *parallax* effects pose the greatest challenge for reliable (high probability of detection, low false alarm rate) detection of moving vehicles and dismounts. In this context, parallax refers to the phenomenon by which stationary clutter objects move through the EO or IR image plane at different rates depending on the distances of the objects from the moving sensor platform. Although a number of 2D image processing techniques have been developed to partially mitigate parallax effects in moving target detection, errors due to parallax can only be optimally removed by using a 3D model to predict the motion of each stationary object from one image to the next, enabling successful separation of moving foreground targets from the stationary background clutter.

In our approach to 3D-based moving target detection, we perform statistical background modeling in the reconstructed triangular mesh 3D surface model framework. That is, we perform texture modeling of the 3D surface using images collected from different perspectives, e.g., the passive 2D EO/IR images from which the model was reconstructed. For moving target detection purposes, texture modeling is performed to produce not just a single intensity per surface location, but rather to produce a statistical model of intensities per surface location. This statistical distribution can be used to produce a texture map of most typical intensities, as we have illustrated in Figure 2. Alternatively, for moving target detection, the statistical background model can be used to predict not only the most probable intensity, but also the expected uncertainty in the intensity (e.g., some materials in the scene, such as vehicle or building windows, produce glint from certain angles, leading to large intensity variations). In comparing the observed pixel intensities with the predicted intensity distributions, change detection can be performed with robustness to typical scene variations, in order to reduce clutter detections. Furthermore, since the statistical intensity modeling is performed in a geo-registered 3D framework, registration of the statistical background model with the current frame is performed with optimal mitigation of parallax effects, including accurate prediction of occlusion/un-occlusion conditions.

In Figure 6, we provide an example of moving vehicle and dismount detection with near-optimal parallax mitigation based on the 3D model shown in Figure 1. In 6(a), a view of the 3D model, in the form of a *depth map*, from the current sensor perspective, is shown. The dark blue in the bottom of the depth map indicates the boundary of the modeled region. The color map indicates *range* from the sensor, with the progression being from *red to blue for shorter to longer range*. In 6(b), a region of interest within the sensor frame is shown. This scenario was selected, for demonstration purposes, to contain severe parallax effects due to the presence of two tall exhaust stacks, numerous buildings, and a large stadium within the field of view. *Automated moving target detections are indicated with red crosshairs drawn by the algorithm on the sensor frame.* The algorithm automatically located the three moving vehicles that an analyst was

able to locate in this scene, as well as an addition dismounted human that the analyst did not notice until cued by the automated detection algorithm.

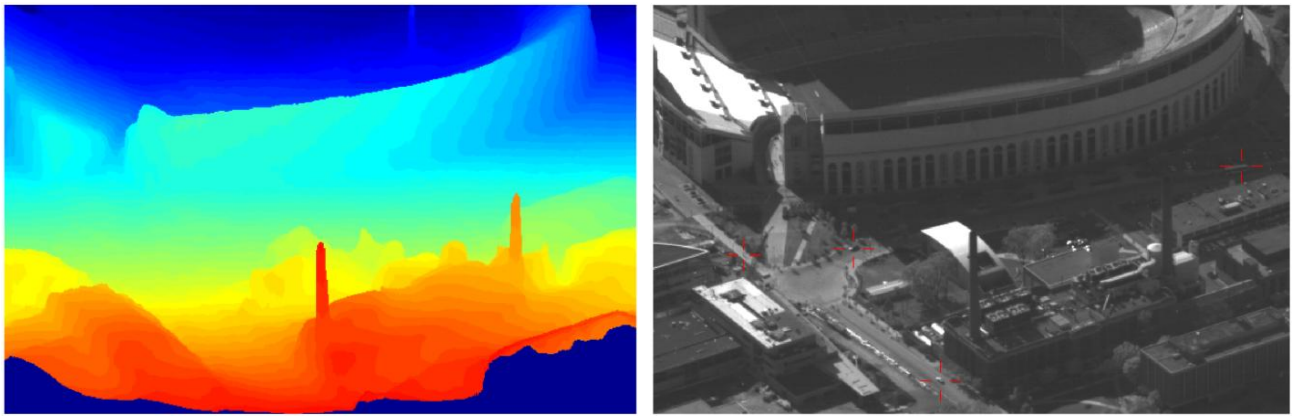


Figure 6. Example automated moving target detection with near-optimal parallax mitigation using an automatically-reconstructed 3D model.

5. CONCLUSIONS

In this paper, we have discussed algorithmic approaches for exploiting wide-area persistent EO/IR motion imagery for multi-sensor geo-registration and automated moving target detection. We first presented enabling capabilities, including sensor auto-calibration and automated high-resolution 3D reconstruction using passive 2D motion imagery. We then presented algorithmic approaches for 3D-based geo-registration, and demonstrated and quantified performance achieved using public release data from AFRL's CLIF 2006 data collection and OGRIP. Finally, we discussed algorithmic approaches for 3D-based moving target detection with near-optimal parallax mitigation, and demonstrated automated detection of dismount and vehicle targets in coarse-resolution CLIF 2006 imagery. Natural extensions of this work include 3D modeling applied to context-aided tracking, e.g., predicting when targets will be occluded, and incorporating user-defined rules for alert generation (e.g., alert for targets scaling structures or positioning on rooftops). Integration with sensor resource management algorithms can also be pursued, e.g., routing small UAVs to observe targets which will soon become occluded by 3D structures. Geo-registered 3D models can also serve as a framework for statistical and syntactic characterization of a scene over time, including modeling of normal behavior and recognizing anomalous behavior.

ACKNOWLEDGEMENTS

The authors would like to thank R. Alan Wood, Todd Rovito, Clark Taylor, Kevin Priddy, and Mark Minardi for their helpful suggestions and support of this work under Air Force contract FA8650-10-C-1709.

REFERENCES

[1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," IEEE Trans. Signal Proc., v. 50, no. 2, Feb., 2002.

[2] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter—Particle Filters for Tracking Applications*, Artech House, 2004.

[3] O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*, MIT Press, 2001.

[4] Curtis Cohenour and Frank van Grass, "Image Georegistration," Final Report Part B, Rev 0, Ohio University, April 25, 2011.

[5] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[6] Andrew P. Brown, "Persistent Electro-Optical / Infra-Red Wide Area Sensor Exploitation," March 2012 Technical Report, Toyon Research Corp.

[7] Joseph L. Mundy and Chung-Fu Chang, "Fusion of Intensity, Texture, and Color in Video Tracking Based on Mutual Information," *AIPR*, 2004:10-15.