Automated UAV-based mapping for airborne reconnaissance and video exploitation

Stephen Se*, Pezhman Firoozfam*, Norman Goldstein*, Linda Wu*, Melanie Dutkiewicz*, Paul Pace**, JLP (Pierre) Naud***

*MDA, 13800 Commerce Parkway, Richmond, B.C., V6V 2J3, Canada
**Defence R&D Canada – Ottawa, 3701 Carling Ave., Ottawa, ON, K1A 0K2, Canada
***Department of National Defence, D Geo Int, Ottawa, ON, K1A 0K2, Canada

{sse,pfiroozf,norman,liwu,mel}@mdacorporation.com
paul.pace@drdc-rddc.gc.ca, pierre.naud@forces.gc.ca

ABSTRACT

Airborne surveillance and reconnaissance are essential for successful military missions. Such capabilities are critical for force protection, situational awareness, mission planning, damage assessment and others. UAVs gather huge amount of video data but it is extremely labour-intensive for operators to analyse hours and hours of received data.

At MDA, we have developed a suite of tools towards automated video exploitation including calibration, visualization, change detection and 3D reconstruction. The on-going work is to improve the robustness of these tools and automate the process as much as possible. Our calibration tool extracts and matches tie-points in the video frames incrementally to recover the camera calibration and poses, which are then refined by bundle adjustment. Our visualization tool stabilizes the video, expands its field-of-view and creates a geo-referenced mosaic from the video frames.

It is important to identify anomalies in a scene, which may include detecting any improvised explosive devices (IED). However, it is tedious and difficult to compare video clips to look for differences manually. Our change detection tool allows the user to load two video clips taken from two passes at different times and flags any changes between them.

3D models are useful for situational awareness, as it is easier to understand the scene by visualizing it in 3D. Our 3D reconstruction tool creates calibrated photo-realistic 3D models from video clips taken from different viewpoints, using both semi-automated and automated approaches. The resulting 3D models also allow distance measurements and line-of-sight analysis.

Keywords: UAV, airborne reconnaissance, video exploitation, mosaicking, change detection, 3D reconstruction, calibration

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are becoming the platform of choice for surveillance operations undertaken by the Canadian forces. UAVs are cheaper than piloted planes, can fly in areas that are dangerous or "dirty" (contaminated), and can undertake dull, repetitive reconnaissance jobs. UAVs can carry many types of sensors; however video cameras are among the most useful since they collect imagery continuously, thus:

- The downlinked data can be used to steer the UAV,
- They are less likely to miss items of interest than non-continuous imaging sensors,
- They can be used to cue other sensors that are non-continuous, and
- They provide frame overlap that allows for 3D reconstruction and mosaicking.

Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VI, edited by Daniel J. Henry,
Proc. of SPIE Vol. 7307, 73070M · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.820225

However, video cameras provide large amounts of data. Using this data efficiently and advantageously is a challenge. How the imagery is used will depend on the mission goals and at which stage of the mission it is used, ranging from using archived imagery for planning to real-time imagery during an operation.

The US Army UAV field manual [1] describes the most common missions performed by UAVs:

- Reconnaissance: Near real-time combat information about terrain, friendly unit actions, and the disposition of possible enemy elements.
- Surveillance: Area surveillance in friendly or enemy territory.
- Situational Awareness: Provide commanders with situational awareness/understanding and mission planning information.
- Security: Reaction time and manoeuvre space for the main body and area security.
- Targeting: Target acquisition, target detection and recognition, target designation and illumination, and battle damage assessment.
- Communication support: Voice and data communications retransmission.
- Movement support: Convoy security, mines/Improvised Explosive Devices (IED) detection.

For the missions described above, data collected from UAVs must in some instances be reviewed quickly, e.g., in support of near real-time operations in the field. However, in other instances, it is analyzed in more depth and over longer timeframes, e.g., to support mission planning.

There are some software packages commercially available that provides video exploitation capabilities. Terrasight [2] from Sarnoff supports situational awareness, targeting, mission planning, and image analysis using full motion video. It addresses the needs for airborne and ground video surveillance, enabling the user to utilize real-time video. The capabilities include stabilization, mosaicking, change detection and geo-registration.

TacitView [3] from 2d3 is a software suite that provides imagery intelligence from aerial video data. It has capabilities including video stabilization, super resolution, mosaicking, moving target indication, change detection, metadata extraction, object segmentation and terrain generation.

Small UAVs are of growing importance for surveillance and reconnaissance tasks, but they only allow a small payload of a few kilograms. Therefore, they typically do not have high precision stabilized sensor platform. Heinze et al [4] have developed a set of image exploitation algorithms for small UAVs including mosaicking, stablilization, image enhancement, and moving target indication. The system is being tested and assessed by military personnel. Gregga et al [5] are building a Real-time Aerial Video Exploitation (RAVE) station for small UAVs. It provides video stabilization, mosaicking, moving target indicators, tracking and target classification.

This paper gives an overview on the suite of tools developed at MDA for video exploitation and the on-going work to make these tools operational and to integrate them with a GIS framework. While most of the video exploitation capabilities have been prototyped in previous projects, the main objectives of the current on-going AMI (Advanced Mapping from Imagery) project is to improve the robustness of these tools, automate the process as much as possible and integrate these tools with ArcGIS. ESRI ArcGIS [6] is a GIS framework with a family of products for visualizing information with geographical locations.

Section 2 will describe the camera calibration which is a critical step required for the other tools. Visualization tools including video stabilization, field-of-view expander and mosaicking will be discussed in Section 3. Section 4 will present the change detection tool while Section 5 will describe the 3D reconstruction tool covering both the semi-automated and automated approaches.

2. CALIBRATION

Camera calibration includes estimating both intrinsic and extrinsic camera parameters. The intrinsic parameters include the focal length, image centre and lens distortion, while the extrinsic parameters refer to the camera position and orientation for each video frame.

Although camera calibration is not a specific goal for the user, it is a critical step that occurs “behind the scenes” and that is required for many of the video exploitation tools. For example, in order to create 3D models or to detect changes, camera parameters for the video frames are needed.

As input video data has corresponding Global Positioning System/Inertial Navigation System (GPS/INS) telemetry, this information can be used to provide the initial estimate for the extrinsic camera parameters. The calibration software will then refine these parameters.

2.1 MV-SLAM

The MV-SLAM (Monocular Visual Simultaneous Localization and Mapping) system was developed at MDA in a previous project to recover camera poses and 3D scenes from monocular video images, based on Andrew Davison’s monocular SLAM work [7][8]. MV-SLAM extracts and matches tie points such as SIFT (Scale Invariant Feature Transform) [9] to compute the camera pose and velocity using the Kalman Filter framework. MV-SLAM simultaneously builds a tie point database and uses this database to localize the camera by matching with new tie points. The output will include a tie point database and the camera poses for each frame.

A sample MV-SLAM result is shown in Figure 1 (left) for a nadir planar poster test data. The camera motion is a circular robot trajectory parallel to a planar poster. The figure shows that the recovered key pose trajectory resembles the circular robot trajectory. 3D features on the poster are also recovered, with the normals more or less pointing perpendicular to the surface of the poster.

As MV-SLAM uses an incremental approach where each image frame is processed one by one, a bundle adjustment step will be carried out to obtain the optimal set of camera parameters for the entire image sequence.

2.2 Bundle Adjustment

Bundle adjustment [10] is an optimization problem on the 3D structure and the viewing parameters (camera pose as well as the intrinsic calibration parameters). From a set of images taken from different viewpoints, the bundle adjuster can simultaneously refine the 3D coordinates describing the scene geometry and the parameters of the relative motion.

In this case, bundle adjustment will take the tie points and camera pose information from MV-SLAM, and perform optimization to refine both the intrinsic and extrinsic calibration parameters. Figure 1 (right) shows an example of bundle adjustment with three frames where the camera poses and a sparse reconstruction of the tie points are recovered. The upper left graph shows the RMS error during the Levenberg-Marquardt non-linear optimization process.

The bundle adjustment will compute a set of accurate camera parameters which can then be used by the change detection and 3D reconstruction tools. For instance, the change detection tool will detect changes between two calibrated video clips and the 3D reconstruction tool will process the calibrated video to generate a photo-realistic 3D model.

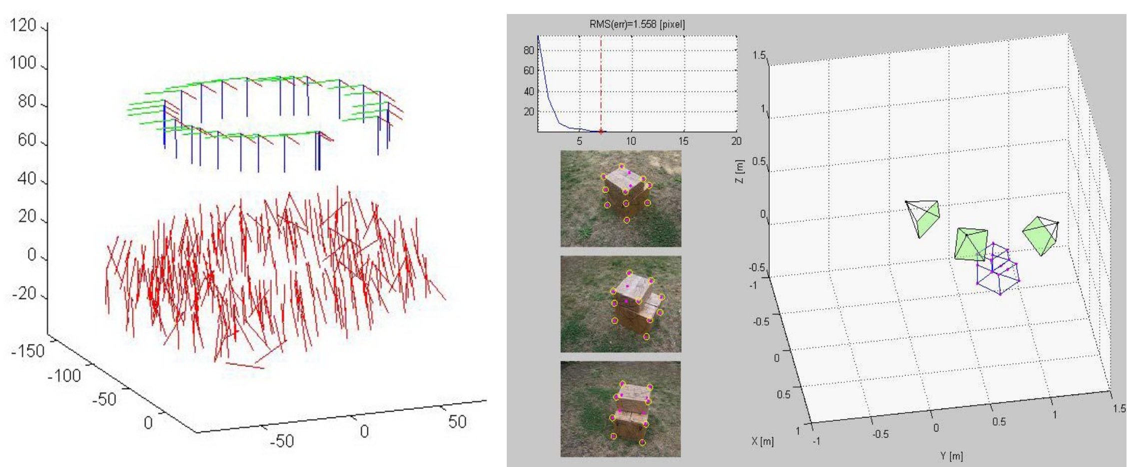

Figure 1 MV-SLAM output including the key pose trajectory and the features (left). Multi-frame bundle adjustment showing the match points in the left three thumbnails and the reconstructed structure on the right (right).

3. VISUALIZATION

3.1 Stabilization

Video stabilization decreases operator fatigue when working with video data. This module allows the operator to stabilize a video stream. Stabilization parameters will be automatically calculated for an input dataset during the calibration phase. The stabilized video can be rendered in real-time from the stabilization parameters. The GUI illustrated in Figure 2 (left) allows the operator to view and record the stabilized video into a new video file. In this example, there is black fill in the upper left corner of the stabilized video. The warping inherent to stabilization will normally produce some black fill in the stabilized product.

3.2 FOV Expansion

The FOV expander module creates a dynamic mosaic which expands the field of the view of current frame from the surrounding frames, hence providing more contexts to the current frame. FOV expander parameters are computed during the calibration phase by aligning the video frames close to the current frame. The GUI illustrated in Figure 2 (right) shows the interface that the operator uses to view and record a new FOV expanded video file.

Figure 2 Video stabilization GUI (left) and FOV expander GUI (right)

3.3 Mosaicking

Video is good at identifying motion, but it is poor at providing good overall situational awareness. It is sometimes desirable to create large still images from the sequence of video frames contained in a video clip.

The mosaicking tool produces a mosaic image from video frames with sufficient overlap between successive frames. The operator will be able to get the big picture over the region of interest. The module will align and blend the video frames into a mosaic within the region of interest. The resulting mosaic will be overlaid on top of a reference map, as shown in Figure 3 (left), to provide a wider context over the area of operation.

4. CHANGE DETECTION

It is important to identify anomalies in a scene, which may include detecting any mines/Improvised Explosive Devices (IED). However, it is tedious and very difficult to compare hours and hours of video clips looking for differences manually. Subtle changes may also be missed by visually comparing two video clips.

The change detection module finds differences between two spatially similar videos (recorded at different times over the same terrain). For the AMI project, we will implement change detection for pairs of video streams that view the ground from similar perspectives. Considering 2D change detection in situations with small parallax differences simplifies the change detection and makes the problem tractable.

Changes between the two video files will be identified and flagged on the reference map according to their locations. The changes will be categorized into three types according to the size of the change, and displayed using green, yellow and red flags to denote small, medium and large changes respectively, as shown in Figure 3 (right). The operator will be able to examine the changes by selecting the flag of interest and the changes would also be exported into a text file to generate reports.

Figure 3 A high resolution mosaic overlaid on a lower resolution reference map (left). Change Detection GUI where the colour flags indicate the different types of changes (right).

5. 3D RECONSTRUCTION

3D models are useful for situational awareness for military reconnaissance, as it is easier to understand the scene by visualizing it in 3D. The 3D reconstruction module will create a photo-realistic 3D model of buildings and vehicles from an input video clip. The 3D model can then be viewed from different perspectives and allows distance measurements and line-of-sight analysis. The quality of the resulting 3D model will depend on the input data, as model creation requires sufficient texture in the video clip for feature tracking, calibration and reconstruction. Both semi-automated and automated approaches will be described for 3D reconstruction.

5.1 Semi-automated 3D reconstruction

Semi-automated 3D reconstruction approaches are usually domain-driven. For example, if the objective is to reconstruct 3D models of buildings interactively, a set of primitives can be provided, such as a particular type of roof or building [11]. When the user selects such a primitive, the system would then estimate the parameters for the primitive such as the height and width of the building based on the image information.

For the AMI project, the user would reconstruct surfaces interactively instead of using 3D primitives. The reconstruction is not restricted to buildings, but it is flexible to allow modeling of other man-made 3D objects such as military vehicles. As each surface would need to be created interactively, this approach is suitable for simple 3D objects that can be modeled with relatively few triangles. For natural or complex 3D objects, such as rough terrain, automated 3D reconstruction would be more suitable, as the terrain would be better modelled by a large number of small triangles.

Figure 4 (left) shows a sample GUI for the 3D reconstruction in semi-automated mode, where the operator will provide correspondences in the video frames. Tools will be provided to assist the operator as much as possible. Figure 4 (right) shows a sample photo-realistic 3D model inside ArcGlobe, where the operator can navigate around to visualize the model from different viewpoints and perform measurements on the model.

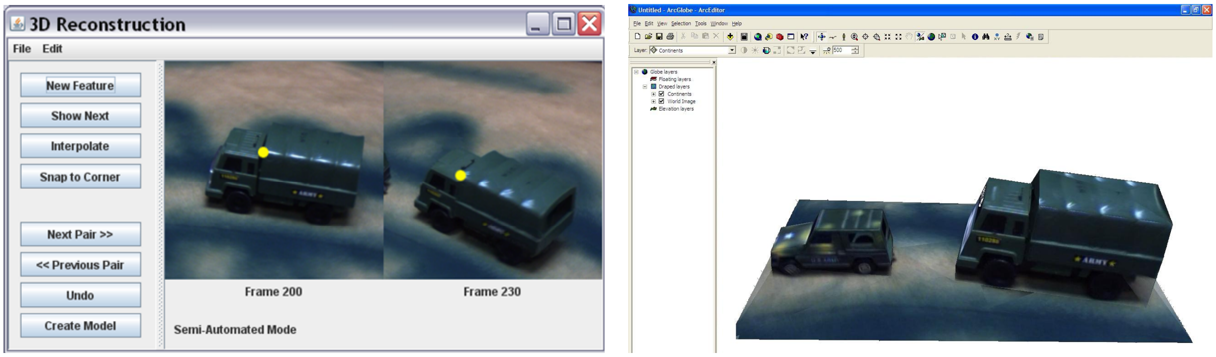

Figure 4 Semi-automated 3D reconstruction GUI (left) and the resulting 3D model visualized in ArcGlobe (right).

5.2 Automated 3D reconstruction

The objective of this tool is to create photo-realistic 3D models from calibrated video sequences automatically. This problem is also called multi-view 3D reconstruction. Multi-view 3D reconstruction algorithms [12] can be categorized into four classes: (a) 3D volumetric approaches, (b) surface evolution techniques, (c) algorithms that compute and merge depth maps, and (d) techniques that grow regions/surfaces from a set of extracted features or seed points.

The automated 3D reconstruction work is still on-going, with the goal of allowing 3D models to be generated without any user input. The proposed approach would process the input frames in a pair-wise fashion, to obtain pair-wise depth maps which are subsequently merged together into a photo-realistic triangular mesh 3D model.

This approach can model natural terrains well but may not reconstruct the sharp edges of man-made objects cleanly. Selected frames in an input video is shown in Figure 5 (left) and a preliminary automated 3D reconstruction result is shown in Figure 5 (right), where calibrated key frames are used to generate a 3D model in KML (Keyhole Markup Language) format. The ground is reconstructed well but some artifacts can be seen at the edges of the boxes.

There is a trade-off between the semi-automated mode and the automated mode. The former approach requires some operator assistance but generates clean 3D models. The automated mode does not require operator assistance but the 3D model may be noisy. The automated mode is more practical than the semi-automated mode for complex scenes that can only be modeled with a large number of triangles, unless the scene can be simplified. Moreover, images need to contain sufficient texture to allow automatic image matching to perform successfully. Depending on the dataset, 3D models generated in the automated mode may require some manual editing by the operator afterwards.

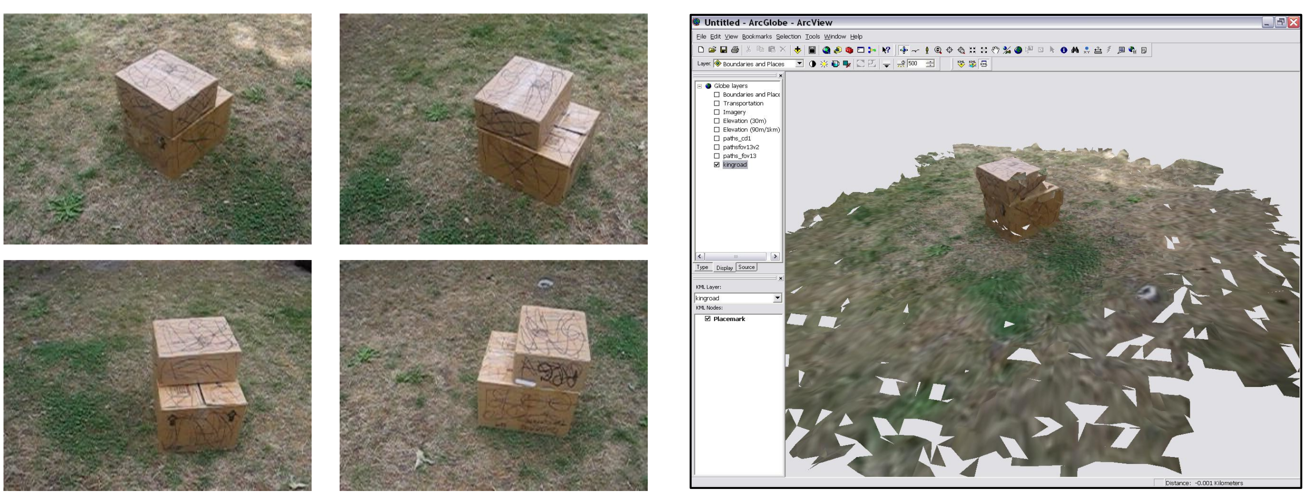

Figure 5 Selected frames in the input video (left) and a preliminary model from automated 3D reconstruction shown in ArcGlobe (right).

6. CONCLUSIONS

Airborne surveillance and reconnaissance are essential for successful military missions. Such capabilities are critical for force protection, situational awareness, mission planning, damage assessment and others. MDA provides UAV services for flying the Heron, which is a medium altitude long endurance UAV system [13]. MDA recently conducted its first UAV flight in Afghanistan in January 2009. On-going flight operations service would support Canadian troops by providing important intelligence, surveillance and reconnaissance information to commanders and front-line soldiers. UAVs gather huge amounts of video data, but it is extremely labour-intensive for operators to analyse such data.

In this paper, we have described a suite of tools developed at MDA towards automated video exploitation including calibration, video stabilization, field-of-view expander, mosaicking, change detection and 3D reconstruction. ArcGIS is being used at both the Canadian military's Mapping and Charting Establishment (MCE) and Canadian Forces Joint Imagery Centre (CFJIC). Therefore, developing on the ArcGIS platform ensures these tools would fit into their current work process.

The on-going work is to integrate these tools with ArcGIS and to make these tools operational so that they are more robust and user-friendly. Calibration tool upgrades and automated 3D reconstruction development are also in progress. The tools developed will be evaluated by Canadian Directorate of Geospatial Intelligence including MCE and CFJIC.

ACKNOWLEDGEMENTS

We would like to thank Dr. Paul Pace, DRDC, for technical guidance, D Geo Int for military support and guidance, in particular Mr. JPL (Pierre) Naud and the DIR program for financial support.

REFERENCES

[1] US Army UAV field manual, http://www.fas.org/irp/doddir/army/fmi3-04-155.pdf

[2] Terrasight, http://www.pyramidvision.com/products/terrasight

[3] TacitView, http://www.2d3.com/product/?v=7

[4] N. Heinze, M. Esswein, W. Kruger and G. Saur, Automatic image exploitation system for small UAVs, Proceedings of SPIE Vol. 6946: Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications V, Orlando, Florida, March 2008.

[5] J.B. Gregga, A. Pope, K. Kielmeyer and Y. Ran, Real-time aerial video exploitation station for small unmanned aerial vehicles, Proceedings of SPIE Vol. 6946: Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications V, Orlando, Florida, March 2008.

[6] ESRI ArcGIS, http://www.esri.com/

[7] A.J. Davison, Real-time simultaneous localization and mapping with a single camera, Proceedings of International Conference on Computer Vision (ICCV), Nice, France, October 2003.

[8] J.M.M. Montiel, J. Civera and A.J. Davison, Unified inverse depth parametrization for monocular SLAM, Proceedings of Robotics : Science and Systems (RSS), Pennsylvania, August 2006.

[9] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60(2):91-110, 2004.

[10] B. Triggs, P. McLauchlan, R. Harley and A. Fitzgibbon, Bundle adjustment – a modern synthesis, Proceedings of the International workshop on Vision Algorithms, 298-372, Corfu, Greece, September 1999.

[11] E. Gulch, H. Muller and T. Labe, Integration of automatic processes into semi-automatic building extraction, International Archives of Photogrammetry and Remote Sensing (IAPRS), Vol. 32, Munich, Germany, 1999.

[12] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, A comparison and evaluation of multi-view stereo reconstruction algorithms, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, June 2006.

[13] MDA Unmanned Systems, http://www.uavs.ca/