Building Aerial Mosaics II: Metadata, Feature Matching, Loop Closure

Esin Turkbeyler, Chris Harris
Roke Manor Research, Romsey, Hampshire SO51 0ZN

Abstract

This paper addresses the task of making a mosaic map with moving track information from images gathered by a downward-looking camera on an airborne platform so that the picture of the ground movement can be exploited by higher-level analysis. In our previous work we successfully showed in simulations that the bundle adjustment technique results in consistent, undistorted maps. In this paper techniques for feature matching between overlapping, but non-sequential, images, where the track information is not available, have been developed in order to apply bundle adjustment on loop closure with real data. These are based on geometric and appearance based feature matching. Moving tracks are also placed on the mosaic image using the homographies of in-between frames and performing local bundle adjustment. Additionally, the bundle adjustment algorithm has been extended to use the NATO STANAG metadata tags in mosaic building. The algorithms are also extended to use a graph decomposition technique in order to provide a scalable solution to real scenarios.

Keywords: UAVs, surveillance, Visual Moving Target Indication, target detection, mosaicing, homography, bundle adjustment, metadata, feature matching.

Introduction

This work aims to develop techniques to detect objects moving on the ground by processing video data from a UAV-mounted camera and then map these detections over a wide area so that the picture of ground movements can be exploited by higher-level analysis. By mapping detected movement over an area, it will be possible to perform a number of higher-level functions such as analysis of movements over longer timescales to establish normal patterns and detect anomalies.

In the first phase of this project, we developed front-end moving target detection algorithms [Ref 1]. In the second phase, we developed and implemented the algorithms for building up the map mosaic with a bespoke simulator [Ref 2]. It has been successfully shown in simulations that the bundle adjustment technique results in consistent, undistorted maps.

In this year's research, we developed techniques to match features between overlapping but non-sequential images (such as between the rows of a raster scan, or on loop closure), in order to apply bundle adjustment techniques to real data. These feature matching techniques are both geometric and appearance based. Moving tracks are also placed on the mosaic image using the homographies of in-between frames and performing local bundle adjustment. Additionally, the bundle adjustment algorithm has been extended to use the NATO STANAG metadata tags in mosaic building. The algorithms are also extended to use graph decomposition techniques in order to provide the scalable solution to real scenarios. These techniques will be described after a brief summary of previous research.

Detection of Movement: In the first phase of this work [Ref 1] algorithms were developed which detect moving objects while discounting the apparent motion of other objects caused by the movement of the camera platform. A three stage algorithm was developed [Ref 1]. The first stage involved calculating the fundamental matrix using a RANSAC-based estimation process. A more robust two stage temporal analysis algorithm was also applied to eliminate outliers.

Mosaic Map: The second phase of the work dealt with mosaic map formation of the video. In our previous research, two main options were identified for mosaic map building of aerial images. These were homography-based methods [Ref 4] and pose-based methods, which require knowledge of pose and a calibrated camera. We chose the homography-based method because it addresses scenarios reported to be of particular interest (deserts, urban terrain on flat ground), and was anticipated to be less computationally complex. Consequently, the sequential chain and bundle adjustment algorithms [Ref 2] developed in the second phase did not require calibration and pose information of the camera. Here, a brief description of the mapping algorithm is given since the work in this paper is built upon this technique.

Mapping Using Homography: Consider a set of images that are to be assembled into a mosaic. Each image will have a so-called absolute homography, one that transforms image coordinates to map coordinates. The mosaicing task is to determine the absolute homographies, using only the measured pair-wise homographies between images. The pair-wise homographies of consecutive homography images are called sequential links [Ref 2], see Figure 1. Other pair-wise homographies can exist, where two homography images have no common tracks, and yet are sufficiently overlapped. Mostly this will occur because the camera view has returned to a similar location after making a sideways excursion as in a raster scan, or after completing a circuit. These pair-wise homographies are called cross-links (see Figure 1).

Bundle Adjustment: Bundle adjustment is only performed when there are cross-links [Ref 2]. When each frame is added to the mosaic, all the constituent frames are adjusted with respect to each other so that the consistency over the entire network is optimised.

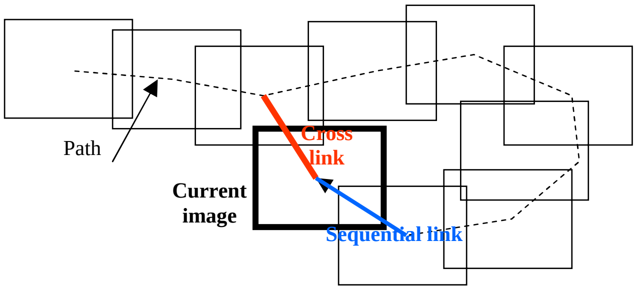

Figure 1 Sequential and cross-link homographies

Feature Matching for Cross-Links

This section describes techniques which address feature matching between overlapping but non-sequential images, where the track information is not available. These features are matched between cross-linked homography frames as in Figure 1. This will enable us to apply bundle adjustment on loop closure with real data when the camera comes back to the same location after making a circuit on a wide area.

Geometric Feature Matching

Geometric feature matching is concerned with whether the image locations of a proposed pair of corresponding features are plausible. Given a feature location in one image, what is the "search" region in another image in which a corresponding feature could lie? To be able to quantify the feature position uncertainties, the uncertainty (or covariance) of the homography parameters must be quantified. To determine the feature search regions, it is necessary to use the homography covariance between a pair of images. By using the appropriate feature

Jacobian matrix provided by bundle adjustment calculations, the homography covariance can be projected down to a pixel covariance, and this rendered at a specified confidence level as an uncertainty ellipse [Ref 2].

Appearance Based Feature Matching

On long path sections without any loop closure events, the homography error can become so large that the geometric search region encompasses the entire frame, and is thus of no discriminatory value. (In other words, we have become lost.) The path lengths over which this happens is some 10-20 fields of view. In these cases, feature matching can only be performed by using the feature appearance.

Appearance-based feature matching concerns making plausible feature or track correspondences, based on local pixel attributes such as shape, colour, and texture. It is important that any appearance measures have good invariance to both geometric transformations that originate from different viewpoints, such as translation, rotation, scale, affine warping, and also to photometric transformations such as contrast.

A feature descriptor with the requisite invariances is the SIFT (Scale Invariant Feature Transform) image descriptor [Ref 5]. Here, these are used to describe the Harris features that comprise the stationary tracks.

To achieve loop closure, the existence of a revisit must be discovered. This is investigated between the latest key frame, and all the previous key frames in which there is geometric uncertainty bigger than one frame. Therefore geometric uncertainty does not apply anymore. SIFT matches are first made, without regard to geometry. Next, RANSAC (RANdom SAmple Consensus) is used to propose homographic transformations, which are assessed by the number of matches that support the proposal. A satisfactory proposal will form a cross-link, which is introduced into bundle adjustment to effect loop closure.

Results - Loop Closure with Real Data

In our real data case, the camera path follows the roads in a loop of length 1km. The route starts and finishes at the roundabout.

Firstly, sequential homographies are used to create the mosaic. Key frames are chosen with 30% overlap of stationary points between frames. In Figure 2, with 42 key frames, the loop is completed at the roundabout. Figure 2 shows the mosaic formed prior to bundle adjustment. By the time the camera has completed the loop and come back to the roundabout, there are no common tracks and geometrical uncertainty is large since the sequential homographies have built up a large error.

Figure 2 Mosaic prior to bundle adjustment

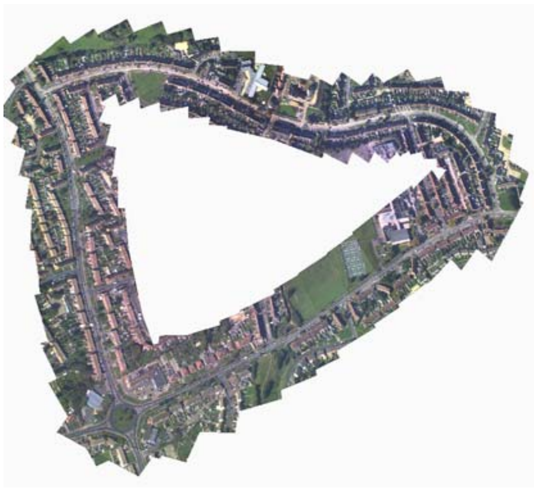

Figure 3 Loop Closure with Bundle Adjustment

In Figure 3, loop closure is achieved with bundle adjustment. Harris corner features are matched using SIFT image descriptors. Features are matched between key frame 1 and key frame 43 and the overlap is detected which results in key frame 43 being cross-linked with key frame 1. The cross-link is then bundle adjusted. Bundle adjustment on the loop closure corrected the errors accumulated through the sequential homography mapping over a large area. Therefore, this is a powerful technique for constructing consistent, undistorted maps over a large area.

Activity Mapping on the Mosaic

Moving tracks which have not been included during mosaic building are placed back into the mosaiced image. Mosaicing and loop closure is performed using key frames only. However, tracks of moving objects will occur on all frames, not just the key frames used for loop closure. To correctly place all of the moving tracks onto the mosaic, the homographies of frames between the key frames are needed. Therefore, we compute the homographies of in-between frames by performing local bundle adjustment with the end-frames fixed at the key frames. This bundle adjustment uses the measured frame pair homographies, and takes linear time so is scalable. As a result the moving tracks are placed on the mosaic map in the right locations without errors, i.e. a moving car is mapped onto the road correctly. This can be clearly seen in Figure 8.

Results - Activity Mapping

The whole activity map is shown in Figure 4. The colours in this figure represent relative average speed of the detected moving tracks – Red indicates the slowest while blue the fastest.

We also obtain the travel direction, and chronological time information from the moving tracks. Therefore we gather the activity over the mosaiced wide area. This information can also be used to understand the moving objects' behaviour.

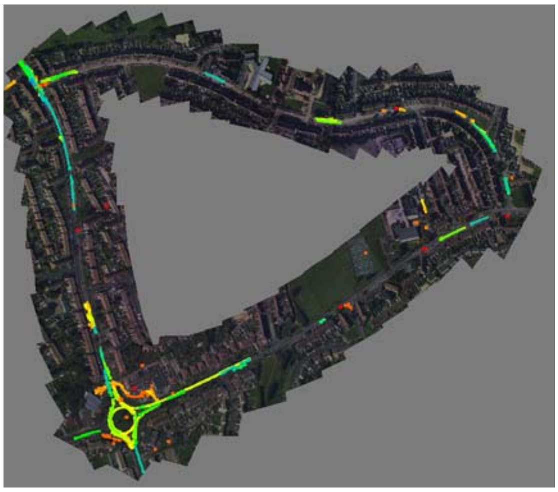

Figure 4 Mosaic image with moving tracks showing relative speed.

When the operator clicks on a moving track on the mosaic map, the original frame from the video recalls the activity of the moving object. If for example the operator clicks on the orange track in Figure 5, a red cross appears on the mosaic and the video frame where the walking person is recalled (as shown in the bottom right corner of Figure 5). In Figure 6, the track clicked by the operator belongs to a van.

Figure 5 Red cross on the mosaic is a person walking on the original video

In Figure 7, red coloured tracks belong to moving objects on the first visit of the aircraft while the blue coloured tracks belong to the moving objects on the second visit of the aircraft. Moreover, the colours in Figure 8 represent the directions of travel.

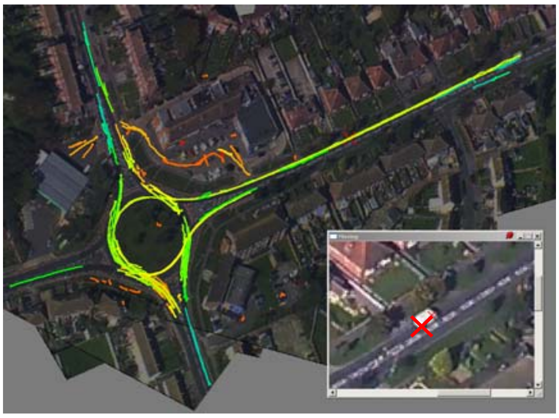

Figure 6 Red cross on the mosaic is a van on the original video

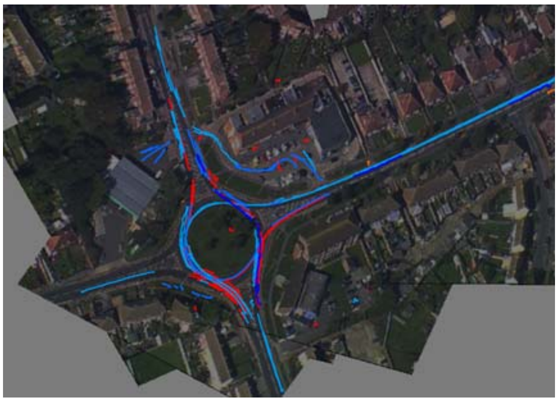

Figure 7 Moving tracks coloured differently in revisit

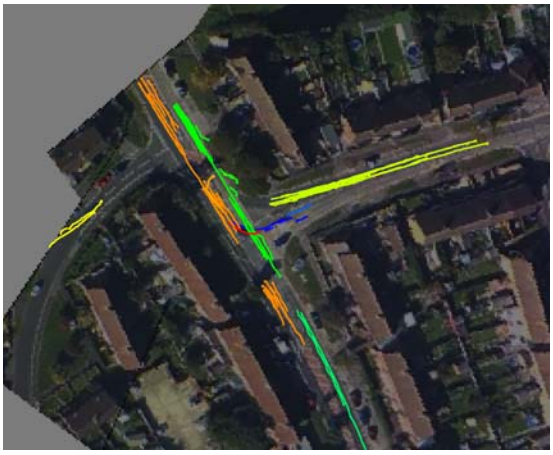

Figure 8 Direction information of tracks

Scalability

With the naïve approach, if there are currently n key frames, the processes have the following complexities:

- Detection of loop closure is O(n), or less if homography covariances can be used;
- Computation of homography covariances is $O(n^3)$;
- Bundle adjustment is $O(n^3)$.

Scalability is increasingly a problem for paths that contain 100 or more key frames because the bundle adjustment algorithm is cubic in the number of key frames. A solution to this problem has been implemented by decomposing the sequential chains of key frames into a graph containing nodes and arcs. The nodes are the key frames contributing to cross-links, but it can also be useful to insert additional nodes to ensure the arcs are not too long. Each arc is treated as a single sequential link, with the homography covariance computed by chaining the constituent homography covariances. This reduces the adjustment algorithm computation to being cubic in the number of nodes rather than number of key frames.

The homography covariance and bundle adjustment will again scale as $O(n^3)$, where n is now the number of nodes. The homography covariance of an arc is obtained by a linear analysis. However, after bundle adjustment, an arc may be strongly flexed, making linear analysis invalid. Therefore, additional nodes can be inserted into long, highly flexed arcs, in order to regain linearity. A simulated example has been used to illustrate the benefits of using graph decomposition methods for scalability.

Figure 9 shows the results of executing a figure-of-eight path of length 75 key frames, starting on the left, and with two loop closures (the crossing point, and start/finish). The nodal images are shown bright, and the non-nodal images are shown in outline. In the background the (darkened) truth is shown, for comparison. In Figure 10, only 4 key frames connected by 3 arcs are used as nodal frames in the bundle adjustment. Almost identical results are obtained. In this example the computational saving in bundle adjustment is a factor of ~ $(75/4)^3$ = 6600 times.

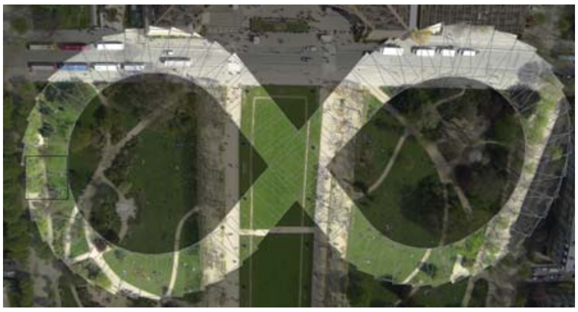

Figure 9 Bundle adjustment using 75 key frames

Figure 10 4 frames as nodal frames used

Metadata

In this paper we also address the integration of NATO STANAG metadata [Ref 6] to our algorithm. The metadata tags used in this paper are positions of four corners of the image. Metadata tags can be used to:

- initialise a mosaic – if initialisation is poor, then a lot of perspective distortion results;
- prevent a mosaic from drifting from the accumulation of errors;
- allow for recovery after a discontinuity, e.g. an interval without sufficient stationary tracks.

The accuracy and the update rate of the metadata tags are two key issues which will affect the accuracy and robustness of the solution. The update rate is not addressed in the NATO STANAG standard [Ref 6] and there could be different rates for different systems. Similarly, accuracy information has not been established.

The metadata tag information has been integrated into the bundle adjustment. The bundle adjustment is a mechanism for minimising measurement residuals, by varying the absolute homographies of the node frames. The metadata tag information has been enrolled into the bundle adjustment simply by including the residuals of the metadata tags at the image locations. An image with metadata tags needs to be designated to be a node frame, because only the node frames are affected by bundle adjustment.

A simulated example has been used to illustrate the benefit of metadata tags. The true path of the image on the ground was a figure-of-eight path in Figure 9.

Figure 11 Incorrect starting homography, no metadata tags

Figure 11 shows the path that result from an incorrect start homography, with a small starting error of some 10% of the image size. The resultant path is, of course, just a homography gauge transformation of that shown in Figure 11, and it has serious errors away from the start.

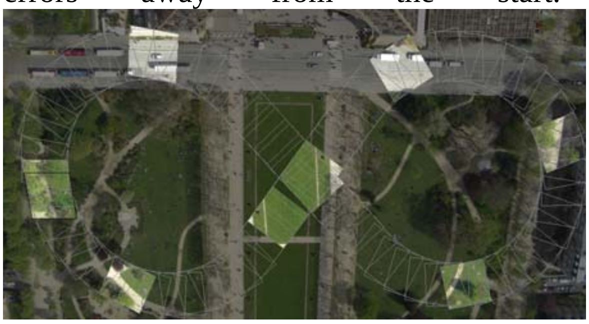

Figure 12Incorrect starting homography with 8 metadata tagged images

The result of the same starting error, but including metadata tag information in each of the nodal images is illustrated in Figure 12. The metadata tags are at the four image corners, and each has an accuracy of the image size. The figure-of-eight path has been completed, with two loop-closing events.

Conclusions

A powerful technique for constructing consistent, undistorted maps with moving track information over a large area has been demonstrated with real airborne data. The mosaic is formed using bundle adjustment techniques with appearance based feature matching for loop closure. The algorithm has been extended to improve scalability with considerable saving on computation time. Additionally, metadata tags information integrated into bundle adjustment has been shown to be useful even if they are only moderately accurate.

Visual MTI together with mosaicing provides the basis of a powerful video analysis aid for airborne videos as indicated in this paper's results.

Future Work

We are in a good position to work on higher level analysis. Statistical analysis of moving features to discover normal and anomalous motion with rule-based techniques could be studied. We could also provide a Visual MTI and mosaicing system as an input to an anomaly detection system (such as VMAD [Ref 3]). The systems similar to VMAD are mainly developed for ground CCTV cameras. The important design criteria in this case will be performance issues and the anticipated relationship between training data volumes, detection thresholds and detection/false alarm probabilities of anomalies.

Acknowledgements

The work reported in this paper was funded by the Electro-magnetic Remote Sensing (EMRS) Defence Technology Centre, established by the UK Ministry of Defence and run by a consortium of SELEX Galileo, Thales UK and Roke Manor Research.

The authors also wish to thank the Police Air Support Unit for assistance in recording the data used in this paper.

References

1 R J Evans, E Turkbeyler "Visual MTI for UAV Systems", 4th EMRS DTC Conference, Edinburgh, July 2007.

2 Esin Turkbeyler, Chris Harris, Richard Evans "Building Aerial Mosaics for Visual MTI", 5th EMRS DTC Conference, Edinburgh, July 2008.

3 R J Evans, E L Brassington "Video Motion Processing for Event Detection and Other Applications" IEE Annual Conference on Visual Image Engineering, VIE2003.

4 F. Caballero, L. Merini, J.Ferruz, A. Ollero, "Homography Based Kalman Filter for Mosaic Building. Application to UAV position estimation", IEEE International Conference on Robotics and Automation, Italy, 10-14 April 2007.

5 David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, January 2004.

6 NATO Motion Imagery STANAG 4609 (Edition 2) Implementation Guide, 2007 http://www.nato.int/docu/stanag/4609/4609_home.htm