

# Building aerial mosaics for visual MTI

This paper addresses the task of making a mosaic from images gathered by a down-looking camera on an airborne platform. This is in the context of a system to detect and map the positions of moving objects. We present three mosaicing approaches based on integrating together sets of measured pairwise homographies, i.e. geometric relationships, between overlapping image frames. The methods are simple chaining, consensus placement and bundle adjustment. We have demonstrated all the approaches with simulated data whilst the simplest way of using pairwise links, simple one-dimensional chaining, has been demonstrated with real data. With the bundle adjustment method, 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. We have successfully shown, in simulation, that the bundle adjustment technique results in much more consistent, undistorted maps.

By

[Esin Turkbeyler](#), [Chris Harris](#) and [Richard Evans](#)

Roke Manor Research, Romsey, Hampshire SO51 0ZN

---

Keywords: UAVs, surveillance, vision processing, Moving Target Indication, target detection, mosaicing, homography, bundle adjustment.

## Introduction

Video cameras have considerable advantages over many other sensors for a number of applications because they are compact, cheap and low power. This paper addresses mosaic map generation for the purpose of autonomous event detection from a camera mounted on a UAV (unmanned air vehicle).

This work builds on earlier research undertaken under the EMRS DTC [3], which considered military applications of an event detection technique called VMAD (Video Motion Anomaly Detection) [2] originally developed for stationary cameras at Roke Manor Research Limited (Roke) for traffic monitoring, security and other civil purposes.

The earlier work extended these techniques to apply them to moving cameras. Two different scenarios of interest have been identified, an emplaced camera (i.e. on pan-tilt mount but set at a fixed location) [3] and a UAV-mounted case where motion in all 6 degrees of freedom was possible [1].

In this paper we build on the work described in [1] to detect moving objects from the camera mounted on the UAV. We aim to map these detections over a wide area so that the picture of ground movements can be exploited by higher-level analysis.

The scenario is a loitering UAV looking downward from a significant height, perhaps flying in a recurrent pattern. It is envisaged that by mapping detected movement over an area it will be possible to perform a number of higher-level functions such as detection of convoys and other coherent movements (rather than simply detecting individual moving objects), and analysis of movements over longer timescales to establish normal patterns and detect exceptions.

The alternative processing techniques will be discussed and the new algorithms to build a mosaic will be described below after a summary of earlier work on front-end Visual MTI [1] to detect moving targets on the ground.

## Visual MTI front-end

In [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 [1] which used in the 1st stage, the fundamental matrix calculated by a RANSAC-based estimation process. (The fundamental matrix encapsulates knowledge of the apparent motion of static features in a scene when observed from different viewpoints.)

1



Front end Visual-MTI results.

The advantage of this approach is that movement can be detected reliably in scenes with significant three-dimensional structure. The 2nd stage of the algorithm investigated temporal analysis by applying the history of feature classifications (i.e. classification as moving or stationary) to provide a more robust algorithm to eliminate the outliers. The 3rd stage attempted to eliminate non-consistency by extending the temporal processing. A typical result from the front-end is shown in Figure 1.

### Alternative processing strategies

We studied options for the mapping stage of processing. This task is essentially equivalent to that of constructing a mosaic from a set of aerial images, though in our case we are not actually concerned with the appearance of the composite image, but instead just the geometric relationships between the positions of objects detected in different frames of an image sequence.

The mosaicing task for a translating camera has been the study of much research, but it is remarkable that there is no generally accepted solution. Two main options were identified:

- Homography-based methods, i.e. methods which assume the ground is essentially flat and as a result do not require the platform position and orientation (pose) to be calculated or instrumented – nor do these methods require a calibrated camera.
- Pose-based methods which require knowledge of pose and a calibrated camera. Pose-based methods such as structure-from-motion or SLAM are mathematically unstable when the camera field of view is small, or the imaged scene is flat-on [4].

As an example of the application of homography-based methods elsewhere [5] deals with the problem of constructing a high quality mosaic of the sea bed. An algorithm is presented for the simultaneous creation of mosaics and the estimation of the camera trajectories using homographies. They have applied a three stage algorithm. The final stage of the algorithm consists of estimating the set of homographies and the world plane description that best fit the observation data. This assumes that the camera calibration is known; thus this technique is a different from our own. They have constructed the circular mosaic image and also recovered 3D camera paths. The camera paths successfully show the closed circuit.

As another example, in [6] a homography-based Kalman filter for mosaic generation is presented. They used a Kalman filter to adjust all homographies between cross-link image pairs.

In this paper we have chosen the homography-based method as the preferred way forward because they address scenarios reported to be of particular interest (deserts, urban terrain on flat ground), and they are anticipated to be less computationally complex.

The new algorithms will be described below after a discussion of homography and the process of constructing the pairwise links.

### Mapping using homography

Mapping concerns assembling a temporal sequence of aerial images, to facilitate the analysis of moving objects detected in the images. The temporal sequence may dwell for an extended period on the same ground location, or may travel over the ground to cover a ground area bigger than covered by the field of view of a single image. The approach chosen was homography. Homography is a relationship between two images, such that there is a one-to-one correspondence between points in each image. For aerial images from a high viewpoint, or of a flat plane, the appropriate geometric form of the correspondences is the perspective transformation, so the homography is called perspective homography.

A perspective transformation is governed by a parametric formula applicable to the whole image, and has 8 parameters. The mapping between a point  $r = (x, y)$  on one image, and a point  $r' = (x', y')$  on another is given by:

#### Equation 1

$$r' = (x', y') = \frac{(a.x + b.y + c, d.x + e.y + f)}{g.x + h.y + 1} \\ \equiv F(r)$$

where  $p = (a..h)$  are the 8 perspective parameters, and  $F$  is the function that transforms  $r$  to  $r'$ . This is an exact transformation for points on a plane that are imaged by an uncalibrated camera from an arbitrary viewpoint.

#### Alternative homography-based mosaicing methods

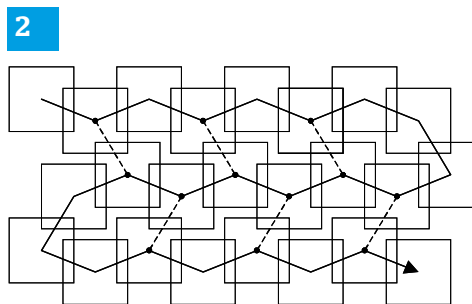
We have identified ways of applying a homography-based method to build a mosaic using pairwise relationships between images of a sequence. These are:

- **Simple Chaining:** Here pairwise links between frames are combined in a one-dimensional chain, reflecting the order of frames in the video sequence. Each new frame is added in sequence to the map (mosaic) in a way which is most consistent with the frame to which it is linked.
- **Consensus Placement:** Here pairwise links are combined in a two-dimensional network, with each new frame added in sequence to the map in a way which is most consistent with frames already in the mosaic to which it is linked.
- **Bundle Adjustment:** Here we again use a two-dimensional network but 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.

#### Calculating pairwise homographies

The first stage is measuring the homography between a pair of images. This requires a number of corresponding features, and so the images must overlap. The features used are Harris corner points. These points can be tracked across a sequence of images, and can be classified as stationary or moving by performing a Fundamental Matrix analysis – as previously reported [1]. It is essential that only the stationary points are used for measuring the pairwise homography. If enough feature tracks run between a pair of images, then they may be used for the measurement of the pairwise homography. The subset of images from the video sequence, which have enough common stationary tracks such that accurate homography can be calculated, are called the homography images.

The residual of Equation 1 is assumed to originate from an independent, unbiased noise process acting on the point positions. By finding the parameters that minimise the noise power over all corresponding points, the maximum likelihood solution is obtained. This involves finding the parameters that minimise the sum of the squared residuals. An iterative algorithm is used, which is initiated using the Direct Linear Transform (where equation (1) is linearised by multiplying by the denominator). This also results in an estimate,  $\Sigma$ , of the covariance of the parameters.



Cross links in image sequence.

### Homography network

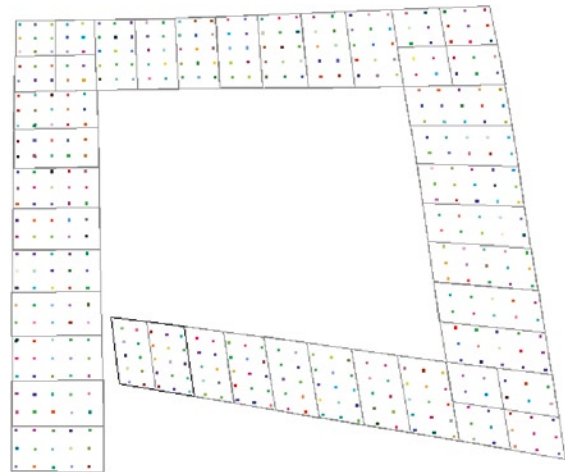
Consider a set of images that are to be assembled into a map (mosaiced). Each image will possess a so-called absolute homography, one that transforms image coordinates to map coordinates. The mosaicing task is to determine the absolute homographies, using only the measured pairwise homographies between images. Due to measurement errors, there will not in general be a solution that exactly agrees with the measured pairwise homographies, and so a best-fitting solution is sought. Note that such a solution will be one of a continuous, equivalent set (a gauge transformation), in which all the absolute homographies are composed with a further arbitrary homography. This would, for example, act to shift, scale, or rotate the map – factors that cannot be resolved from the imagery alone.

The pairwise homographies of consecutive homography images are called sequential links. Other measurable pairwise homographies can exist, where two homography images have no (or too few) 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, see Figure 2. These pairwise homographies are called cross links.

### Simple chain of homographies

The simplest homography network is a simple linear chain, where each homography image is positioned using only its previous neighbour. In this case, the absolute homography of each image is optimally determined using only the sequential link to the previous image and the maximum likelihood solution is obtained. Each absolute homography will remain unaltered by subsequent images. However, map distortions will increase indefinitely with the length of the chain.

3



Square loop with simple chain.

Simple chaining is illustrated in Figure 3 for simulated data. Here the camera looks straight down and executes an exact square circuit over the ground, starting at bottom left. The mapping of each image into the ground is shown as a grey quadrilateral and the features with the image as coloured dots. Each image has a 50% overlap with the previous image. A simulated measurement noise of 0.2 pixels causes the homography parameters to become increasingly incorrect for each successive image, and results in a distorted map.

### Simple chain of homographies – example with real imagery

Homography mapping has been undertaken on real imagery for the roundabout sequence, as shown in Figure 4. In this sequence, the plane flew a complete circuit around a motorway roundabout. The camera view was some  $60^\circ$  from vertical. Note that the scene is not strongly planar (heights span at least 10 metres). Processing has only been by simple chaining of homographies. In the initial map, the angular shape seen is due to the masking out of annotations. The initial absolute homography has been chosen manually to represent a vertical viewpoint. The homography of each subsequent image is determined by simple chaining, and is placed on the map simply overwriting any previous contents. The lack of cross-linking results in some drift (random walk) of the homography parameters by the end of the sequence, causing map misalignment.

### Consensus replacement

Cross-links can be used to reduce the distortions. The simplest method of using cross-links is to position the current image using the sequential link, together with any cross-links. This is called consensus placement, and forces the current image into a compromise position. Consensus placement for a raster scan is illustrated in Figure 5. The scan starts at bottom left, and successively sweeps up and down the map while moving to the right. The undistorted map shape is a square. The measurement noise was 1 pixel.

### Bundle adjustment

An improvement to consensus placement is to vary the absolute homography of every image, so as to minimise the residuals of both the sequential and cross links. This is called bundle adjustment. Homography covariances have been computed and used to obtain the maximum likelihood solution.

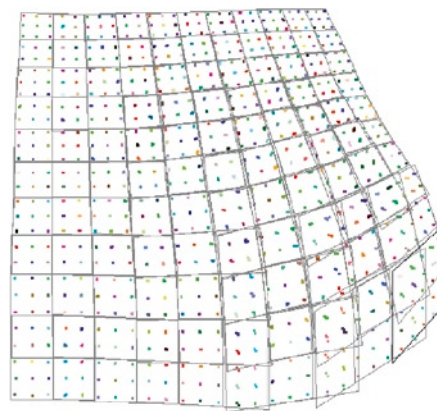
Bundle adjustment is an iterative scheme, each step of which performs a Gauss-Newton minimisation over all the homography parameters. Not all the parameters may be allowed to vary, because invariance of the links to a gauge transformation makes the matrix singular, and so the first image has been set to remain fixed. In Figure 6 are shown the first and last (converged) iterations. Note that very little distortion remains. Usually 3–5 iterations are sufficient to achieve convergence.

4



Mosaic forming with simple chaining using real data.

5



Raster scan with consensus replacement.

### Bundle adjustment example with simulated images

The absolute homographies can be used to transform the image content, and accumulate them into the map. This is illustrated for a raster scan in Figure 6, before and after bundle adjustment. Unlike previous figures, the simulated corresponding points have not been displayed, even though they have still been used for calculations. The image content here does not affect the processing – as feature extraction and matching (for sequential links and cross links) have been simulated. The individual simulated image frames have been generated by sampling a satellite image taken from the web.

### Homography covariance

Homography covariances have been computed and used to obtain the maximum likelihood solution. Homography covariances can additionally be used to search for feature correspondences for cross-links. To determine feature search regions, it is necessary to use the homography covariance between a pair of images. By using the appropriate Jacobian, the homography covariance can be projected down to a pixel covariance, and this is rendered at a specified confidence level as an ellipse. This is illustrated in Figure 7, for a square path both before and after loop closure. Each ellipse shows the positional uncertainty of the centre of a past image, with respect to the current image (outlined in black). Putting the final image in place greatly reduces the homography covariances, and hence the size of the search regions.

6a



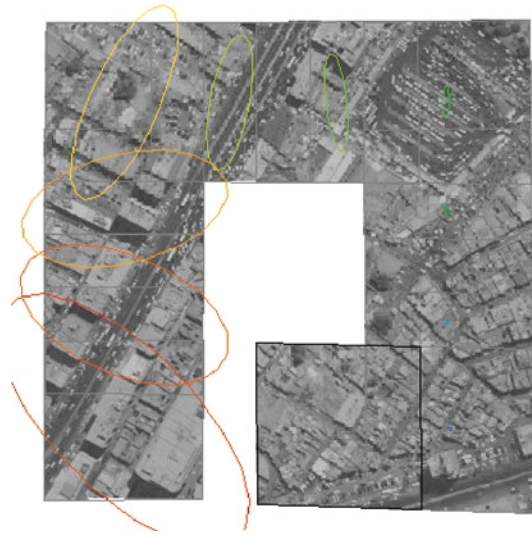
Before bundle adjustment.

6b



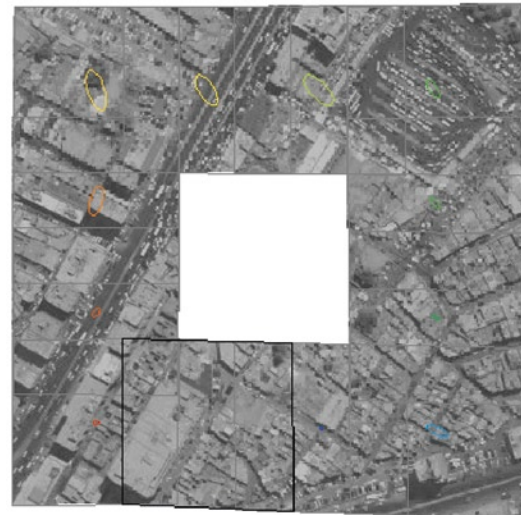
After bundle adjustment. Using homography for image content mapping – a) before and b) after bundle adjustment.

7a



Before loop closure.

7b



After loop closure Homography covariances.

### Conclusions

Perspective homography mapping from aerial images has been designed and implemented. Processing of real data has been limited to simple chaining of homographies. This technique has been seen to work, though with some level of map distortion. The bundle adjustment technique has been successfully tested on synthetic data. This promises to be a powerful technique for constructing consistent, undistorted maps over a large area. The key elements of the processing chain, to map the positions of moving objects over an area, have been developed.

### Future work

We have successfully shown in simulation that the bundle adjustment technique results in much more consistent, undistorted maps. To apply this technique to real data however, we will develop techniques to match features between overlapping but nonsequential images (such as between the rows of a raster scan). This will complete the processing chain so that a full ground movement picture can be extracted and made available for higher level analysis.

In a naïve implementation, bundle adjustment is computationally intensive and approaches to improving its scalability will be implemented.

### 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, Roke Manor Research and Filtronic.

### References

1. R J Evans, E Turkbeyler 'Visual MTI for UAV Systems', 4th EMRS DTC Conference, Edinburgh, July 2007.
2. R J Evans, E L Brassington 'Video Motion Processing for Event Detection and Other Applications', IEE Annual Conference on Visual Image Engineering, VIE2003, University of Surrey.
3. R J Evans, R G Porges 'Video Motion Anomaly Detection for Military Applications', 3rd EMRS DTC Technical Conference, Edinburgh 2006.
4. R J Evans, P Saddington, C. G. Harris, '3D Computer Vision – Unpublished Report'.
5. F. Caballero, L. Merini, J.Ferruz, A. Ollero, 'Homography Based Kalman Filter for Mosaic Building. Application to UAV position estimation', 2007 IEEE International Conference on Robotics and Automation Roma, Italy, 10–14 April 2007.
6. N. R. E. Gracias, 'Mosaic-based Visual Navigation for Autonomous Underwater Vehicles', PhD Thesis, 2002.