The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Creation of Geo-Referenced Mosaics from MAV Video and Telemetry Using Constrained Optimization and Bundle Adjustment

Benjamin Heiner, Clark N. Taylor
MAGICC Lab
Brigham Young University, Provo UT

Abstract— **Miniature Aerial Vehicles (MAVs) are quickly gaining acceptance as a platform for performing remote sensing. However, because MAVs are flown close to the ground (300 meters or less in altitude), their field of view for any one image is relatively small. In addition, the context of the video (where and at what orientation are the objects being observed, the relationship between images) is unclear from any one image. To overcome these problems, we propose a geo-referenced mosaicing method that creates a mosaic from the captured images and geo-references the mosaic using information from the MAV IMU/GPS unit. Our method utilizes bundle adjustment within a constrained optimization framework. Using real MAV video, we have demonstrated our mosaic creation process on over 700 frames. Our method has been shown to produce the high quality mosaics to within 7m using tightly synchronized MAV telemetry data and to within 30m using only GPS information (i.e. no roll and pitch information).**

Index Terms— **Mosaic, Bundle Adjustment, MAV, constrained optimization**

I. Introduction

In recent years, civilian and military agencies have increased their utilization of miniature unmanned aerial vehicles (MAVs - less than a 2m wingspan) in many information gathering missions, including rural search and rescue, agricultural information gathering, and reconnaissance and surveillance. Due to their small size, MAVs are attractive platforms for executing these missions. MAVs possess several advantages: they can be deployed quickly and repeatedly, their small size simplifies storage and reduces their detectability, they have lower costs when compared to larger UAVs and manned aircraft, they enable operators to explore hazardous environments without risk of life, and they can obtain imagery at sub-decimeter resolution due to their low flight altitude.

While sub-decimeter resolution imagery is easily obtained through flight of a MAV, presenting the data to the end-user in a format amenable to analysis is complicated due to several issues. First, during the flight of an MAV, hundreds of pictures can be collected by the MAV that must be analyzed by the user. For example, a 10 minute flight – collecting 1 image per second – will require the user to analyze 600 images. Second, even when the user is looking at an image, the "context" (geo-location, which direction is North, and relation of the image to other images) of each image is not immediately apparent. In order to overcome these problems, a single, large, integrated image (mosaic) can be created from the image sequence. This mosaic allows the user to quickly analyze all of the visual data collected and reveals the spatial relationship between images. The creation of mosaics from multiple images is a well-studied problem [1], [2], [3], [4]

This work funded by AFOSR award number FA9550-07-1-0167

Fig. 1: This image sequence demonstrates the noisy pose estimations in MAV Video Sequences. Note how inaccuracies in the pose estimates cause discontinuities in the global mosaic.

In addition to creating a mosaic, however, it is useful to geo-reference (determine the GPS location of every pixel) the mosaic to provide more context to the captured images. One method of producing the geo-referencing information for the mosaic is to use the initial pose estimates, from the Inertial Measurement Unit (IMU) and the Global Positioning Systems (GPS), to project the MAV imagery onto a common coordinate system. When coupled with a high precision camera and reference imagery, this method can produce highly accurate mosaics [5]. This method, however, is not well suited to MAVs. The small size of an MAV airframe limits the weight, size, and power available for payloads, necessitating the use of low-quality sensors for the IMU, leading to noisy pose estimates. The effects of a noisy pose estimation can be seen in Figure 1. In this figure, several images are placed in a mosaic using only the information from the IMU and GPS unit. As shown, the placement of the images varies significantly with noise in the pose estimates, causing discontinuities between consecutive video frames. Other techniques which create a continuous mosaic, but geo-reference from the pose data of a single frame ([6], [7]) similarly suffer from poor performance due to inaccurate pose estimates from the IMU/GPS system on-board the MAV.

In order to overcome the effects of noisy pose estimates, [8], [9] creates a consistent frame-to-frame mosaic from the

video stream and then registers the mosaic to a preexisting georeferenced image. In [10], a homography is computed for each frame in a video sequence using frame-to-frame and frame-to-map registration. However, these methods are dependent on the availability of reference imagery, a significant assumption. In addition, when reference imagery is available for the region of interest, it may be out of date or have significant differences in appearance due to differing environmental conditions (i.e. lighting, structures, season of the year, etc.), resolution, or imaging technologies (i.e. IR, EO, etc.) between the current image and the geo-referenced image, significantly impairing the image-to-map registration process.

An alternate approach to building the mosaic is to simultaneously estimate the camera poses and georeferenced feature locations from overlapping images in the video sequence, a process know as Bundle Adjustment (BA) [11]. In a general sense, this process functions by identifying a set of parameters to be estimated, and a set of measured values which are to be modeled as nonlinear functions of those parameters. It then begins a iterative process that estimates the parameter values that predict the measured values most correctly according to some cost function. Most BA methods use the location of salient feature points as the parameter space to be estimated. This approach minimizes the reprojection error between the image locations of several observed and predicted image points and has been shown to produce high quality mosaics [12], [13]. However, these processes have a high computational cost associated with the solving of their normal equations and do not typically make use of geographic information obtained from any IMU/GPS systems attached to the camera.

In this paper, we present a novel modification to traditional bundle adjustment (BA) methods. If the scene is planar (a universal assumption among all aerial mosaicing applications), there is a projective mapping between any two images of the plane. This mapping is commonly referred to as a homography [14]. Rather than using feature points that are tracked over a set of images, we assume that frame-to-frame homographies have been estimated and will accurately align frames with overlap.

The key insight used to set up our constrained optimization problem is that the frame-to-frame homography mapping is also a function of the poses of the camera at the time the images were taken. Because they define a "visually appealing" mosaic, the computed mosaics can be treated as constraints on the true pose estimates. A measurement of this pose is directly computed by the IMU/GPS system on-board the MAV. If we assume that for each frame an estimate of the camera location is returned by the IMU/GPS system, then a constrained optimization routine can be used to determine the set of camera poses that are closest to the measured pose values while meeting the constraints imposed by the frame-to-frame homographies.

This approach has a number of advantages. First, traditional bundle adjustment requires a parameter space that is of size $3M+7N$, where M is the number of features throughout the entire video that were tracked, and N is the number of images being used to create the mosaic. Our method's parameter vector is of size $7N+7$, a significant reduction in size. Similarly, the measurement space in a traditional BA is $2k_iM$ where k_i is the number of images in which feature i appears. Because $M > N$, our method's measurement space of size $8(N-1)+7N+7$ is significantly smaller than traditional BA. This reduction in the size of the measurement and parameter space reduces the computational cost associated with solving the BA normal equations. Second, because the homographies are used as a measurement rather than feature locations, the removal of outliers which is performed to compute the homography significantly reduces the probably that outliers will be included in the BA process. Third, by including the global location information that is available from the MAV's IMU/GPS system in our BA, the most probable geo-location of the mosaic, as a function of pose parameters, is explicitly computed in our constrained optimization framework.

The remainder of the paper is organized as follows. In Section II, describe our algorithm for creating accurately geo-referenced mosaics. In Section III, results demonstrating the abilities of our system are shown. Section IV concludes the paper.

II. Process Overview

In this section we describe our proposed system to generate a single, large, integrated mosaic $\mathbf{M}$ from multiple smaller images ($\mathbf{I}_{i=0\rightarrow n}$), retrieved from a continuous video sequence and their associated pose estimates. These inputs are represented by the left-most box in Figure 2. The pose estimates,$\hat{P}_i$, of the MAV contain the position T and attitude $\vec{q}$ information of the body. The position is encoded using a local North-East-Down (NED) Cartesian coordinate system where north axis is x, the east axis y, and the down axis z. In this coordinate system positive altitudes translate into a negative z. The attitude information is represented as a quaternion [15], [16], providing a simple way to meet the unity norm constraint when used within an optimization algorithm. Therefore, each pose can be represented as follows

$$\vec{P} = \begin{bmatrix} x & y & z & \vec{q} \end{bmatrix}^T. \quad (1)$$

The computational process utilizing these inputs is subdivided into three main steps as shown in Figure 2. Each of these steps are described in more detail below.

A. Pre-Processing

The pre-processing step starts with the removal of the lens distortions from the camera. We assume that the camera has been calibrated a-priori, making the lens distortion a known quantity. The image is then warped to remove lens distortion using function calls from OpenCV [17]. We then generate the frame-to-frame transformations $\hat{\mathbf{H}}_{i\leftarrow i+1}$ for each consecutive image pair in the MAV video sequence. To compute the transformation matrix, $\hat{\mathbf{H}}_{i\leftarrow i+1}$, we first find features in the first image using a Harris corner detector [18]. These features are then tracked in the second image using the pyramidal Lucas-Kanade method [19]. Once feature correspondences

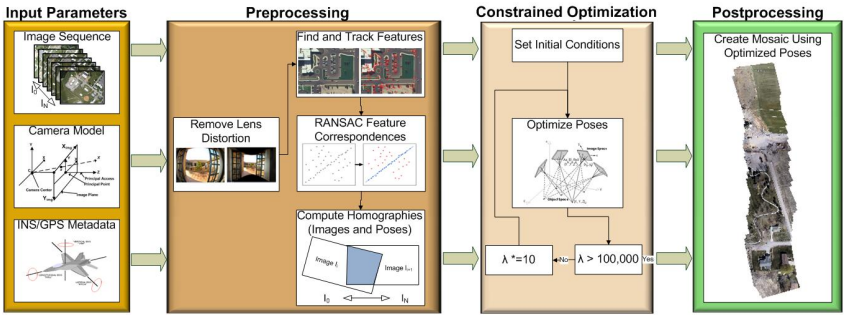


Fig. 2: A flow diagram of our system for creating geo-referenced mosaics from images captured by an MAV.

are found, they are refined using the RANSAC [20] and LMS outlier rejection[21] on the inliers output by RANSAC. After removing outlier correspondences, the transformation matrix $\hat{\mathbf{H}}_{i\rightarrow i+1}$ is computed using the singular value decomposition (SVD) as described in [14]. These homographies are supplied to the constrained optimization procedure described in the following section.

B. Constrained Optimization Using Bundle Adjustment

As discussed in the introduction, the fundamental approach used to create a geo-referenced mosaic is to perform a constrained optimization problem using bundle adjustment. The optimization problem we are trying to solve is:

$$\begin{aligned} \min_{\vec{P}_i,\vec{n},\vec{q}_b} \sum_{i=1}^{N} (\vec{P}_i - \hat{P}_i)^2 \quad s.t. \\ \vec{h}(\vec{P}_i,\vec{P}_{i+1},\vec{n},\vec{q}_b) = \hat{h}_i, \\ ||\vec{n}||_2 = 1, \\ \text{and } ||\vec{q}_b|| = 1 \end{aligned} \tag{2}$$

where $\vec{n}$ is the normal vector (in world coordinates) of the plane being imaged, $\vec{q}_b$ is a "bias" quaternion representing constant errors in the pose estimates from the IMU/GPS unit on-board the MAV, $\vec{h}(\vec{P}_i,\vec{P}_{i+1})$ is a function, described in Section II-C, for computing the homography given two poses and the normal vector, and $\hat{h}_i$ is derived from the matrix $\hat{\mathbf{H}}_{i\leftarrow i+1}$. Because the matrix $\hat{\mathbf{H}}$ is defined up to a scale factor, the bottom right element of this matrix is set to 1. $\hat{h}_i$ is then set equal to

$$\hat{h}_i = \begin{bmatrix} \hat{\mathbf{H}}_{1,1} & \hat{\mathbf{H}}_{1,2} & \hat{\mathbf{H}}_{1,3} & \hat{\mathbf{H}}_{2,1} & \hat{\mathbf{H}}_{2,2} & \hat{\mathbf{H}}_{2,3} & \hat{\mathbf{H}}_{3,1} & \hat{\mathbf{H}}_{3,2} \end{bmatrix}^T \tag{3}$$

where the subindices #,# represent the row and column, respectively, of $\hat{\mathbf{H}}$.

We handle the constraints imposed on this optimization in two ways. The constraints on $\vec{n}$ and $\vec{q}_b$ are imposed by re-normalizing these vectors to one after each iteration of the bundle adjustment [15]. In addition, the poses each contain a quaternion within them representing an attitude which are also normalized to 1. The constraint on $\vec{h}(\vec{P}_i,\vec{P}_{i+1},\vec{n},\vec{q}_b)$ is handled using a Lagrange multiplier λ, leading to

$$\begin{aligned} \min_{\vec{P}_i,\vec{n},\vec{q}_b} \sum_{i=1}^{N} (\vec{P}_i - \hat{P}_i)^2 + \lambda \left(\vec{h}(\vec{P}_i,\vec{P}_{i+1}) - \hat{h}_i\right)^2 s.t. \\ ||\vec{q}_b||_2 = 1 \text{and } ||\vec{n}||_2 = 1. \end{aligned}$$

For a given value of λ, this minimization function can be solved using bundle adjustment. Generally we start with a λ of .001, increasing by factors of 10 until $\lambda = 100,000$. For non-final values of λ, we do not enforce as strict of a convergence criteria for bundle adjustment in order to speed up the optimization process.

For initial conditions on our constrained optimization problem, we initialize the estimated poses to the measured poses returned by the IMU/GPS system on-board the MAV. The normal vector $\vec{n}$ is initialized to a vertical vector ($\begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T$), and the bias quaternion to the identity quaternion ($\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$).

In the following two subsections, we describe the two remaining portions of our constrained optimization problem. First, we describe the function $\vec{h}(\vec{P}_i,\vec{P}_{i+1},\vec{n},\vec{q}_b)$ which is used to enforce the constraints imposed by the computed homographies. Second, we discuss the setup of the bundle adjustment and its relationship to the chosen λ.

C. Computing $\vec{h}(\vec{P}_i,\vec{P}_{i+1},\vec{n},\vec{q}_b)$

As discussed in the prior section, to enforce the constraints imposed by the computed homographies on the estimated poses of the MAV camera, it is necessary to have a function that maps the current pose estimates into a homography matrix. While this function can be derived using pre-existing techniques, we describe it in more detail here to provide a complete description of our system.

To derive the matrix $\mathbf{H}_{i\leftarrow j}$ from world poses, we require the pose estimates $\vec{P}_i$ and $\vec{P}_j$. Because we are also estimating a bias quaternion, the first step of computing $\mathbf{H}$ is to compose the quaternions in the pose estimates with the bias quaternion. These modified poses are used throughout the rest of this section without explicitly stating that the attitude estimates have been biased.

From rigid body motion, we can define the relation between the camera coordinate frame f_c and the world frame f_w as

$$\vec{X}_c = \mathbf{R}_{w\rightarrow c}(\vec{q})\vec{X}_w + T_c, \tag{4}$$

where $\vec{X}_c = [x,y,z]^T$ is a point in f_c, $R_{w\rightarrow c}$ is a rotation matrix (defined as a function of $\vec{q}$ in the pose estimate), and T_c is the location of the origin of f_c if f_w. Solving for $\vec{X}_w$ in terms

of f_{cj} and substituting it into equation 4, we obtain

$$\vec{X}_{ci} = \mathbf{R}_{w\to ci}\left(\mathbf{R}_{w\leftarrow cj}\vec{X}_{cj} + T_{cj} - T_{ci}\right), \quad (5)$$

the rigid body transformation from f_{cj} to f_{ci}. Note that this equation, however, is general for all points in f_{cj}. We will now add the constraint that all points we are interested on lie on a plane.

Let π be a planar surface where all $\vec{X} \in \pi$ and $\vec{n}$ is a normalized vector orthogonal to π. The existence of π implies that

$$\left\langle \mathbf{R}_{w\to c}\vec{n}, \vec{X}_c \right\rangle = d, \quad \frac{1}{d}\left(\mathbf{R}_{w\to c}\vec{n}\right)^T \vec{X}_c d = 1, \quad (6)$$

for all $\vec{X} \in \pi$, where $d = z_c$ is the minimal distance of the camera from the plane. We can now substitute equation 6 into equation 5 yielding

$$\vec{X}_{ci} = \mathbf{R}_{w\to ci}\left(\mathbf{R}_{w\leftarrow cj}\vec{X}_{cj} + T_{ji}\frac{1}{z_{cj}}\left(\mathbf{R}_{w\to c}\vec{n}\right)^T \vec{X}_{cj}\right), \quad (7)$$

where $T_{ji} = T_{cj} - T_{ci}$. Simplifying and adding in the calibration matrices K_{ci} and K_{cj} for the two cameras, we can now rewrite the equation for the perspective transformation matrix $H_{i\leftarrow j}$ as

$$\mathbf{H}_{i\leftarrow j} = \mathbf{K}_{ci}\mathbf{R}_{w\to ci}\left(I + T_{ji}\frac{1}{z_{cj}}\vec{n}^T\right)\mathbf{R}_{w\leftarrow cj}\mathbf{K}_{cj}^{-1}. \quad (8)$$

D. Bundle Adjustment Implementation

Bundle adjustment is a method for finding the closest point on a manifold defined by a parameter vector $\vec{p}$ to a measurement vector $\vec{x}$. In our case, the measurement vector can be written as:

$$\hat{x} = [\hat{P}_0, \cdots, \hat{P}_N, \hat{h}_0, \cdots, \hat{h}_{n-1}, \hat{n}_v, \hat{q}_0]^T, \quad (9)$$

where $\hat{n}_v = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T$ and $\hat{q}_0 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$ are added to the measurement vector to represent prior knowledge of typical values for $\vec{n}$ and $\vec{q}_b$ (the ground should have a normal that is close to vertical, and the biases should be small.)

The parameter space over which bundle adjustment will iterate is defined by the parameters which are being optimized in the constrained optimization problem (P_i,$\vec{n}$, and$\vec{q}_{bias}$). The total parameter vector is defined as:

$$\vec{p} = [\vec{P}_0, \cdots, \vec{P}_N, \vec{n}, \vec{q}_{bias}]^T. \quad (10)$$

To perform this optimization we use a weighted Gauss-Newton method. This method iteratively linearizes the function to be minimized in the neighborhood of the current estimate, by solving linear systems known as *normal equations*. The normal equations are defined as

$$\mathscr{A} = \mathbf{J}^T\Sigma^{-1}\mathbf{J} \quad (11)$$

$$\vec{g} = \mathbf{J}^T\Sigma^{-1}\vec{\varepsilon}_p \quad (12)$$

$$\vec{x}_E^+ = \mathscr{A}^{-1}\vec{g} \quad (13)$$

where $\mathbf{J}$ is an $m \times n$ matrix containing the partial derivatives of the cost function (i.e. a Jacobian matrix) , Σ is the covariance matrix which represents the distance metrics which are defined below, $\mathscr{A}$ is an $m \times m$ matrix that contains the approximated second derivatives of the cost function, $\vec{\varepsilon}_P$ is the residual error in pose estimates and homography matrices, $\vec{g}$ is the gradient, and change $\vec{x}_E^+$ in the parameters for the next iteration. The Jacobian Matrix is formulated as

$$\mathbf{J} = \begin{bmatrix} \mathscr{P} \\ \mathscr{H} \end{bmatrix}, \quad (14)$$

where

$$\mathscr{P} = diag\left(\frac{\delta\vec{P}_1}{\delta\vec{P}_1}, \frac{\delta\vec{P}_2}{\delta\vec{P}_2}, \cdots, \frac{\delta\vec{P}_n}{\delta\vec{P}_n}, \frac{\delta\vec{n}}{\delta\vec{n}}, \frac{\delta\vec{q}_{bias}}{\delta\vec{q}_{bias}}\right) = I_{n+7}, \quad (15)$$

and

$$\mathscr{H} = diag\left(\begin{bmatrix} \frac{\delta\mathbf{H}_1}{\delta P_1} & \frac{\delta\mathbf{H}_1}{\delta P_2} \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\delta\mathbf{H}_{n-1}}{\delta P_{n-1}} & \frac{\delta\mathbf{H}_{n-1}}{\delta P_n} \end{bmatrix}\right), \quad (16)$$

Due to the block diagonal nature of both $\mathscr{P}$ and $\mathscr{H}$, efficient bundle adjustment can be used to solve the normal equations.

The covariance matrix Σ is assumed to be diagonal with weighting for each element as defined in Table I. The weighting on the x,y, z, and $\vec{q}$ parameters represent the assumed error present in IMU/GPS estimates. (Due to the quality of the sensors used in IMU system, the covariance on the attitude parameter q is relatively large.) The weighting of $\vec{n}$ is chosen such that it only allows minimal deviations from the initial estimate of $\vec{n}_v$. The covariance for $\vec{q}_b$ is also chosen to be quite large to allow for a large range of bias values (more motivation for these values will be found in the results section).

While the covariance values described above all have fairly straight-forward physical meanings, the covariance on the homography values is a bit more complicated. The goal of the constraint in the constrained optimization is to ensure that the poses chosen for the MAV camera cause the images to be perfectly aligned. Therefore, we derive the covariances to represent movement in pixel locations due to changes in the homography parameters. Because different elements of the homography matrix have different effects on the pixels, we have covariances that vary as shown in Table I. To utilize bundle adjustment within a constrained optimization framework, the homography covariances are also all scaled by $\frac{1}{\lambda}$.

TABLE I: Covariance Matrix Weights

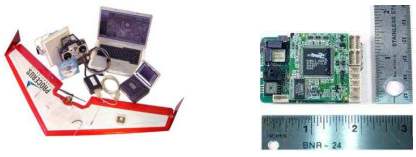
Parameters	Weighting
x, y, z	1
$\vec{q}$	.1
$\vec{n}$	.0001
$\vec{q}_b$	1
$\mathbf{H}_{1,1}\mathbf{H}_{1,2}\mathbf{H}_{2,1}\mathbf{H}_{2,2}$	$\frac{1}{\lambda\ max\left(I_{height}, I_{width}\right)}$
$\mathbf{H}_{1,3}\mathbf{H}_{2,3}$	$\frac{1}{\lambda}$
$\mathbf{H}_{3,1}\mathbf{H}_{3,2}$	$\frac{1}{\lambda\ max\left(I_{height}, I_{width}\right)^2}$

E. Post-Processing

Upon completing the pose optimization process as described above, a single, large, integrated georeferenced mosaic of the region of interest is created using the optimized pose estimates $\vec{P}_{opt,i=0\rightarrow n}$. This is done via the use of a virtual camera centered one meter above the region of interest. Using the virtual camera and optimized poses $\vec{P}_{opt,i=0\rightarrow n}$, we project the MAV imagery onto a global coordinate system. An alpha blend with $\alpha = 0.5$ was used to combined images together with the mosaic.

III. Results

In order to evaluate our system, we used two different sets of images. The first data set was collected from a hand launchable Delta-wing test platform constructed from EPP foam as shown in Figure 3(a). This platform has a wingspan of ˜1.5m, an empty weight of 3.2lbs, and has a payload capacity of 1.8lbs. It is equipped with a 640x480 SONY camera and 2.4GHz NTSC video transmitter (30 frames a second). The pose estimates for each frame are generated by the autopilot controlling the airframe, a Kestrel autopilot by Procerus, shown in Figure 3(b)). This autopilot contains three-axis accelerometers and gyros and two pressure sensors for air speed and altitude measurements and is linked to a GPS estimate to generate full pose estimates. The pose estimated by the Kestrel is transmitted to the ground at 25Hz over a 115.2 kBaud radio modem. This platform is advantageous because Procerus' software enables frame-level synchronization of pose information (including attitude) from the autopilot with video collected by the MAV.



(a) (b)

Fig. 3: Example of Procerus autopilot-based system used during evaluation. (a) D-Wing airframe (b) Kestrel Autopilot.

The second image set was collected using an alternate test platform. This test platform is an MAV used by the US Armed Forces and has a wingspan of 1.3 m and weighs about 4.2 lbs. The platform was equipped with a Canon SD1000 Elph 7 Mpixel commercial camera, which was programmed to take one high-resolution image per second. Imagery and pose estimates collected from the MAV are synchronized using the EXIF time stamps of the imagery and GPS time stamps of pose estimates. Note that tight synchronization between this airframe's autopilot data and imagery was *not* achieved, leading to erroneous attitude estimates (i.e. no roll and pitch information). GPS, however, with its slower update rate, was effectively synchronized with the image data.

Using the first image sequence, we demonstrated the effectiveness of our method by creating a mosaic that consists of over 700 frames of data spanning an area over 500 meters long. The mosaic created from this sequence can be seen in Figures 4 and 5. Notice how vehicles, curbs, roads, buildings, and even rocks in the mosaic are crisp and clear. In addition, notice how lines are well defined and straight. By visual inspection, we can see that the process creates a clear and consistent mosaic. Futhermore, the geo-location of objects within the image appears to be very accurate. Using Google Earth, we identified several objects in the image and measured the distance between the mosaiced image and the Google Earth image of the same object. The geo-location errors were less than 7m across the entire mosaic. Given the size of the MAV airframe and the type of sensors used in the IMU these are promising results.

Using the second image sequence was more difficult. Because the alternate test platform does not return telemetry data which is tightly synchronized with the video, we found the attitude estimates in the telemetry file to be of little use. Therefore, the estimated pose for each image was simply a camera pointing straight down. In this case, we created a mosaic from 30 frames of data spanning an area over 1500 meters long. The mosaic created from this sequence can be seen in Figures 4 and 5. By visual inspection of location at the ends and center of the mosaic, we found that geo-location errors for this mosaic ranged from 16 to 30 meters. This is a significant improvement compared with the raw telemetry information, which led to errors in the range of 9 to 320 meters. These geo-location errors seem promising considering the lack of attitude information about the camera poses.

IV. Conclusions

We have presented a novel modification to the traditional BA method enabling the creation of geo-referenced mosaics from from MAV video and telemetry. This method makes direct use of pose information obtained from an IMU/GPS unit thereby allowing us to accurately geo-reference the mosaic to the world.

The key insight used to set up this bundle adjustment problem is that the frame-to-frame homography mapping is also a function of the poses of the camera at the time the images were taken. Because they define a "visually appealing" mosaic, the computed perspective mappings can be treated as constraints on the true pose estimates. A measurement of this pose is directly computed by the IMU/GPS system on-board the MAV. If we assume that for each frame an estimate of the camera location is returned by the IMU/GPS system, then a constrained optimization routine can be used to determine the set of camera poses that are closest to the measured pose values while meeting the constraints imposed by the frame-to-frame homographies.

Using our method, we have demonstrated mosaics created from over 400 images resulting in geo-location errors of less than 7m. We have also demonstrated geo-registration without any attitude information from the MAV. Visually appealing mosaics were achieved in both cases.

In the future, we would like to extend this method by adding in non-temporal frame-to-frame registration (i.e. topology inference). We will also investigate extending this method to work with non-planar terrain, more significant color variations between images due to either environmental factors or imaging technologies, and super resolution.

REFERENCES

[1] M. Hansen, P. Anandan, K. Dana, G. van der Wal, P. Burt, D. Center, and N. Princeton, "Real-time scene stabilization and mosaic construction," *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pp. 54–62, 1994.

[2] M. Irani, P. Anandan, and S. Hsu, "Mosaic based representations of video sequences and their applications," in *International Conference on Computer Vision*, 1995.

[3] F. Dufaux and F. Moscheni, "Background mosaicking for low bit rate video coding," in *In Proc. ICIP*, Lausanne, Switzerland, 1996.

[4] R. Szeliski and H.-Y. Shum, "Creating full view panoramic image mosaics and environment maps," in *SIGGRAPH*, 1997, pp. 251–258.

[5] D. S. A. Brown, "High accuracy autonomous image georeferencing using a gps/inertial-aided digital imaging system," *Institute of Navigation*, January 2002.

[6] S. Negahdaripour and X. Xu, "Mosaic-based positioning and improved motion-estimation methods for automatic navigation of submersible vehicles," *IEEE Journal of Oceanic Engineering*, vol. 27, pp. 79–99, 2002.

[7] F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "Improving vision-based planar motion estimation for unammaned aerial vehicles through online mosaicing," in *2006 IEEE International Conference on Robotics and Automation*, 2006.

[8] R. Kumar, H. Sawhney, J. C. Asmuth, A. Pope, and S. Hsu, "Registration of video to geo-referenced imagery," in *Fourteenth International Conference on Pattern Recognition*, vol. 2, Aug 1998, pp. 1393–1400.

[9] R. Kumar, S. Samarasekera, S. Hsu, and K. Hanna, "Registration of highly-oblique and zoomed in aerial video to reference imagery," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 4, Sep 2000, pp. 303–307.

[10] Y. L. G. Medioni, "Map-enhanced uav image sequence registration and synchronization of multiple image sequences," *Computer Vision and Pattern Recognition*, pp. 1–7, 2007.

[11] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment–a modern synthesis," *Vision Algorithms: Theory and Practice*, vol. 1883, pp. 298–372, 2000.

[12] H. S. S. Steve Hsu and R. Kumar, "Automated mosaics via topology inference," *IEEE Computer Graphics and Applications*, vol. 22, pp. 44–54, Mar/Apr 2002.

[13] O. P. H. Singh, "Toward large-area mosaicing for underwater scientific applications," *IEE Journal of Oceanic Engineering*, vol. 28, no. 4, oct 2003.

[14] R. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge, UK: Cambridge University Press, 2003.

[15] J. Diebel, "Representing attitude: Euler angles, unit quaternions, and rotation vectors," vol. 1-35, October 2006.

[16] J. J. Kuffner, "Effective sampling and distance metrics for 3d rigid body path planning," in *Proc. IEEE International Conference on Robotics and Automation ICRA '04*, vol. 4, Apr 26–May 1, 2004, pp. 3993–3998.

[17] *Intel Integrated Performance Primitives Reference Manual*, A70805-021us ed., Intel, 2007.

[18] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.

[19] J.-Y. Bouguet, "Pyramidal implementation of the Lucas Kanade feature tracker," Intel Corporation, Microprocessor Research Labs, Tech. Rep., 2000.

[20] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[21] Å. Björck, *Numerical Methods for Least Squares Problems*. Philadelphia: SIAM, 1996.

Fig. 4: This figure shows the quality of an actual MAV mosaic created using 600 MAV images of Vineyard in Utah.

Fig. 5: Geo-referenced mosaic of Vineyard in Utah created using the optimized pose estimates. The geo-locations in this mosaic are 7m or less.

Fig. 6: This figure shows the quality of an actual MAV mosaic created using high-resolution images of Florida location..

Fig. 7: Geo-referenced Mosaic demonstrating the robustness of our optimization method using only GPS data from the autopilot.