Autonomous Helicopter Tracking and Localization Using a Self-Surveying Camera Array

Masayoshi Matsuoka, Alan Chen, Surya P. N. Singh, Adam Coates, Andrew Y. Ng, and Sebastian Thrun

Stanford University; Stanford, CA 94305
{m.matsuoka,aychen,spns,acoates,ayn,thrun}@stanford.edu

Summary. This paper describes an algorithm that tracks and localizes a helicopter using a ground-based trinocular camera array. The three cameras are placed independently in an arbitrary arrangement that allows each camera to view the helicopter's flight volume. The helicopter then flies an unplanned path that allows the cameras to self-survey utilizing an algorithm based on structure from motion and bundle adjustment. This yields the camera's extrinsic parameters allowing for real-time positioning of the helicopter's position in a camera array based coordinate frame. In fielded experiments, there is less than a $2m$ RMS tracking error and the update rate of $20Hz$ is comparable to DGPS update rates. This system has successfully been integrated with an IMU to provide a positioning system for autonomous hovering.

Keywords: structure from motion, bundle adjustment, self-surveying cameras, camera tracking, camera localization

1 Introduction

Position estimation is of critical importance in autonomous robotics research as it is the principal measurement used in machine control and localizing collected data. [1] We utilize three ground based cameras to track and localize one of the Stanford autonomous helicopters (Fig. 1). This system replaces an onboard DGPS system, making the positioning system more robust during aggressive flight maneuvers. DGPS is unreliable because directional GPS antennas are prone to signal occlusions during rolls and omnidirectional antennas are susceptible to multipath during upright flight. Also, by moving the positioning equipment off the helicopter, the weight is reduced allowing the helicopter more power for maneuvering. The cameras are placed on the ground in unsurveyed positions that will allow them to see the helicopter at all times. Because the rotation and translation relationship between each camera is unknown, this extrinsic data will need to be extracted through self-surveying of

P. Corke and S. Sukkarieh (Eds.): Field and Service Robotics, STAR 25, pp. 19–30, 2006.

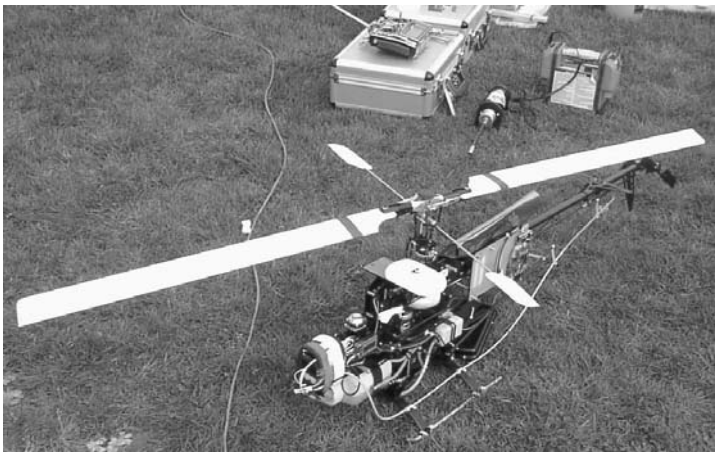


Fig. 1. One of Stanford's autonomous helicopters

the array. Once the extrinsic data has been determined, then the 3-D location of the aerial vehicle can be accurately and robustly tracked in a camera array based coordinate frame with standard least squares, LS, techniques.

The core problems in this project are the localization of the helicopter in each image frame and the self-surveying of the extrinsic parameters for the three cameras. Background differencing is used to locate the helicopter in each image. Essentially, by identifying the background through an average of previous scenes the moving helicopter can be identified as cluster of points in the foreground image. The center of this cluster identifies the approximate center of the helicopter.

Extrinsic information is usually obtained via calibration of the cameras in the scene utilizing a calibration object, such as a cube with a checkerboard pattern, or the cameras are fixed in locations and orientations with known extrinsic parameters. [2] This is not ideal in a field environment because the above methods would require a recalibration of the cameras with a large calibration aid every time a camera is jostled or would require a large structure that would fix the cameras in relation to each other while providing enough coverage to view the entire scene. Thus, the process of camera self-surveying is crucial to the tracking problem. Through this, the camera array will be able to estimate its geometry on the fly while deployed in the field without requiring modifications to the scene or the helicopter. Our approach uses multiple observations of the same scene motion to recover the extrinsic relationships between the cameras. In particular, this is done using a variant of the structure from motion (SFM) algorithm [3] and bundle adjustment [4].

Surveying and calibration will be used interchangeably throughout this paper. When we talk about calibration however, we are only referring to calibrating the extrinsic parameters of our camera array.

2 Background

There are several related localization approaches in the field. [5, 6, 7] Approaches like DGPS [5, 6] and radar provide high precision localization accuracy, but tend to be expensive, hard to relocate, prone to occlusion, or have to be deployed on the vehicle. Like the directional GPS antenna, an on-board camera system is also susceptible to occlusions when the helicopter rolls and pitches.[7] Inertial techniques provide high fidelity, but introduce significant drift error. Our system can be useful as a low-cost portable alternative to standard positioning systems without adding hardware to the helicopter.

The self-surveying ability of our system allows us to place the cameras anywhere on a field such that the cameras cover the operating space where the helicopter will fly and have the helicopter in focus. Self-calibration to acquire extrinsic parameters has been done by groups in the past. [8, 9] The main difference is that they move the stereo cameras in order to extract parameters while we will be moving a point in the image to extract the same type of information. For example, Knight and Reid use a stereo head that rotates around an axis to give calibration and head geometry. [10] Zhang shows that four points and several images from a stereo pair which has moved randomly, but is constant with respect to each other, can be used to compute the relative location and orientation of the cameras along with the 3-D structure of the points up to a scale factor. [11] Our self-surveying technique utilizes an algorithm developed by Poelman and Kanade. They use one camera tracking several feature points and take a stream of images while moving the camera. With this data, they can determine the motion of the camera and the coordinates of each of the feature points. [3]

3 Tracking Approach

A background differencing method is utilized to extract the location of the helicopter in images coordinates from the black and white pictures. First, the statistical model of the background is built by updating a running average of the image sequence over time, with I as a pixel intensity value:

$$I_j^{background}(u,v) = (1-\alpha)I_{j-1}^{background}(u,v) + \alpha I_j^{current}(u,v) \tag{1}$$

where α regulates updating speed. [4] Next, the algorithm takes an image difference of the current image and the background image, and then thresholds out the image difference caused by noise:

$$I_j^{difference}(u,v) = \{ \begin{matrix} I = I_j^{current}(u,v) - I_j^{background}(u,v) & |I \geq I^{threshold} \\ 0 & |I < I^{threshold} \end{matrix} \} \tag{2}$$

Finally, the estimate of a moving object in the image coordinate (u_j, v_j) is estimated by the population mean of the non-zero pixel distribution of the image difference:

$$u_j = \frac{1}{k} \sum_m \sum_n m I_j^{difference}(m,n) \tag{3}$$

$$v_j = \frac{1}{k} \sum_m \sum_n n I_j^{difference}(m,n) \tag{4}$$

Here, the search window (m, n) is a square mask, containing k pixels, centered at the helicopter location in the previous time step. This eliminates unrealistic abrupt jumps in the helicopter location estimate caused by noise and other moving objects elsewhere in the image.

This simple windowed background differencing method works when the helicopter is the principal actively moving object in the search window. Although slow-moving disturbances like clouds in the sky can be distinguished from the helicopter by tuning α and the threshold to appropriate values, this algorithm may be confused when other fast moving objects are in its windowed view, such as swaying trees or airplanes in the background.

As suggested in related literature, the tracking performance can be greatly improved by taking the probabilities of the predicted target dynamics into consideration, for instance, using Kalman filtering [2], the condensation algorithm [12], or multiple hypothesis tracking [13]. In this research, the Kalman filter approach is implemented to improve robustness in maintaining a lock on the helicopter in this specific helicopter tracking environment. However, in the experimental setup used in section 5.3 (in which the helicopter flies above the treeline in each of the camera views), the algorithm does well even without a Kalman filter.

4 Self-Calibration Algorithm

4.1 Structure from Motion

To calibrate the extrinsic parameters of the system, a structure from motion technique based on the algorithm defined by Poelman and Kanade in 1997 will be used for an initial estimate. [3] As opposed to taking a single camera and taking a stream of images of an object as we move the camera, we will use static cameras and take a stream of images as we move the object in the scene. This will provide the data necessary to utilize the algorithm described below.

The equation below shows the standard camera conversion equations:

$$p_j = R_i(P_j + t_i) \tag{5}$$

$$R_i = \begin{pmatrix} i_i \\ j_i \\ k_i \end{pmatrix},\ t_i = \begin{pmatrix} t_{ix} \\ t_{iy} \\ t_{iz} \end{pmatrix},\ p_j = \begin{pmatrix} p_{jx} \\ p_{jy} \\ p_{jz} \end{pmatrix},\ P_j = \begin{pmatrix} P_{jx} \\ P_{jy} \\ P_{jz} \end{pmatrix} \tag{6}$$

M : number of cameras (3)
N : length of flight
i : camera (1,2,...,M)
j : sampling epoch (1,2,...,N)
t_i : the location of the camera i in the world frame
P_j : the helicopter trajectory in the world frame
p_{ij} : the helicopter trajectory in camera i frame
R_i : rotation matrix for camera i
u_{ij}, v_{ij} : pixel values of the helicopter at epoch j in camera i

To convert from 3-D camera frame coordinates to a 2-D image frame coordinate system, a scaled orthographic projection, also know as "weak perspective," will be used. This projection technique, shown in the equation below, approximates perspective projections when the object in the image is near the image center and does not vary a large amount in the axis perpendicular to the camera's image plane. The equations below assume unit focal length and that the world's origin is now fixed at the center of mass of the objects in view.

$$x_i = \frac{t_i \cdot i_i}{z_i},\ y_i = \frac{t_i \cdot j_i}{z_i},\ z_i = t_i \cdot k_i \tag{7}$$

$$u_{ij} = \frac{p_{jx}}{z_i} = m_i \cdot P_j + x_i \tag{8}$$

$$v_{ij} = \frac{p_{jy}}{z_i} = n_i \cdot P_j + y_i \tag{9}$$

$$m_i = \frac{i_i}{z_i},\ n_i = \frac{j_i}{z_i} \tag{10}$$

$$W = R^* P + t^* \tag{11}$$

$$W = \begin{pmatrix} u_{11} & \dots & u_{1N} \\ v_{11} & \dots & v_{1N} \\ \vdots & & \vdots \\ u_{M1} & \dots & u_{MN} \\ v_{M1} & \dots & v_{MN} \end{pmatrix},\ R^* = \begin{pmatrix} m_1 \\ n_1 \\ \vdots \\ m_M \\ n_M \end{pmatrix},\ t^* = \begin{pmatrix} x_1 & \dots & x_1 \\ y_1 & \dots & y_1 \\ \vdots & & \vdots \\ x_M & \dots & x_M \\ y_M & \dots & y_M \end{pmatrix} \tag{12}$$

Using the helicopter's trajectory in each of the cameras, (u_{ij}, v_{ij}), we can solve for the measurement matrix W^*. Taking the singular value decomposition of W^* and ignoring any right or left singular eigenvectors that correspond with the 4th or higher singular values (that appear due to noise) results with:

$$x_i = \frac{1}{N}\sum_{j=1}^{N} u_{ij},\ y_i = \frac{1}{N}\sum_{j=1}^{N} v_{ij} \tag{13}$$

$$W^* = W - t^* = R^* P \approx U_{2M\times 3}\Sigma_{3\times 3}V^T_{3\times N} = \tilde{R}\tilde{P} \tag{14}$$

$$\tilde{R} = U,\ \tilde{P} = \Sigma V^T \tag{15}$$

$\tilde{R}$ and $\tilde{P}$ represent the affine camera positions and the affine structure of the points in the scene respectively which can then be transferred back to Euclidian space with a matrix Q. To determine Q we will use the $2M + 1$ linear constraints defined below. The last constraint will avoid the trivial solution satisfied by everything being zero.

$$W^* = \tilde{R}QQ^{-1}\tilde{P} \tag{16}$$

$$|m_i|^2 = |n_i|^2 = \frac{1}{z_i^2} \Rightarrow |m_i| - |n_i| = 0 \tag{17}$$

$$m_i \cdot n_i = 0 \tag{18}$$

$$|m_1| = 1 \tag{19}$$

With these constraints and the Jacobi Transformation of Q the affine system can then be converted back into Euclidian space. If the resulting Q is not positive definite, then distortions, possibly due to noise, perspective effects, insufficient rotation in the system, or a planar flight path, has overcome the third singular value of W. [3]

We multiply all the rotation matrices and the newly found matrix of points by R_1^{-1} to convert everything into a coordinate frame based on the camera 1 image frame.

After this process, the only remaining extrinsic parameters still unknown is t_i . To find t_i, LS can be used by expanding the equation below to encompass all the points in each camera.

$$\begin{pmatrix} u_{ij} \\ v_{ij} \\ z_i \end{pmatrix} - \begin{pmatrix} i_i \cdot p_j \\ j_i \cdot p_j \\ 0 \end{pmatrix} = R_i t_i \tag{20}$$

The minimum number of points required to self-survey with structure from motion is defined by

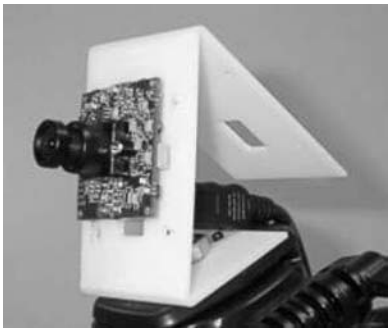
$$2MN > 8M + 3N - 12 \tag{21}$$

Given that three cameras will be used, a minimum of four points will be necessary to self-survey. Because our cameras are static, we can fly the helicopter to four different locations and record images at each location. This will provide the minimum points necessary to self-survey. [14]

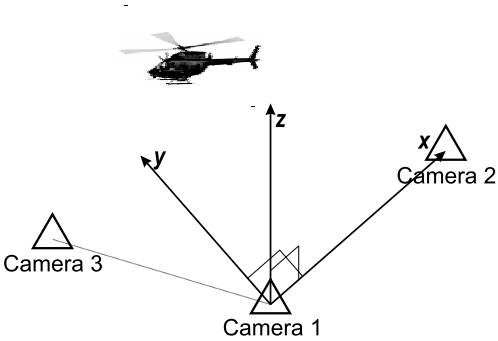
4.2 Camera Frame to World Frame

The resulting extrinsic parameters of the camera array are unscaled and given in the camera 1 image frame. To extract the unknown scale factor inherent to these type of vision problems, the distance L between camera 1 and camera 2 is measured. The ratio of that distance to the unscaled distance between camera 1 and camera 2 is defined as the scale factor.

To rotate the extrinsic parameters from the camera 1 image frame to a world frame, a rotation matrix is created based on the following assumptions (see also Fig. 2(b)):



(a) Point Grey Firefly2 camera.



(b) World frame

Fig. 2. Camera setup

1. Camera 1 is at the origin of the world frame.
2. The vector from camera 1 to camera 2 is the x axis.
3. All the cameras are in the $x - y$ plane.
4. The y axis is defined as towards the helicopter, but orthogonal to the x axis and in the $x - y$ plane.
5. The z axis is then defined by the right hand rule (approximately straight up)

This results in:

$$t_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, t_2 = \begin{bmatrix} L \\ 0 \\ 0 \end{bmatrix}, t_3 = \begin{bmatrix} x_3 \\ y_3 \\ 0 \end{bmatrix} \tag{22}$$

4.3 Bundle Adjustment

Given the SFM solution as initial estimate, the calibration parameters can be refined further by solving nonlinear perspective equations directly via iterative LS, bundle adjustment. [4] The bundle adjustment technique optimizes the calibration parameters, exploring the best array geometry that matches to the set of visual tracking measurements collected during a calibration flight.

The calibration parameters estimated by the LS batch process include the camera locations in the world, the camera orientations, and the helicopter trajectory. Specifically, the following extrinsic parameters are the unknowns to be estimated: t_i, P_j, and the Euler angles associated with R_i, (α_i, β_i, and γ_i).

The set of normalized 2-D tracking points, (u_{ij}, v_{ij}), in the image coordinates is the sole measurement used in this calibration process (except for the measurement L). The following perspective geometry equations relate all

the unknown parameters to the 2-D tracking points via a nonlinear perspective model. [2] (23) is different than (8) and (9) because here we are using a perspective model for the cameras.

$$u_{ij} = \frac{P_{jx}}{P_{jz}} \ , \quad v_{ij} = \frac{P_{jy}}{P_{jz}} \tag{23}$$

$$R_i = \begin{bmatrix} \cos\gamma_i & \sin\gamma_i & 0 \\ -\sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta_i & 0 & -\sin\beta_i \\ 0 & 1 & 0 \\ \sin\beta_i & 0 & \cos\beta_i \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_i & \sin\alpha_i \\ 0 & -\sin\alpha_i & \cos\alpha_i \end{bmatrix} \tag{24}$$

The bundle adjustment method linearizes the perspective equations (5), (23), and (24) into a Jacobian form, and then batch-estimates the unknown calibration parameters via the iterative LS by taking the pseudo-inverse of the Jacobian matrix J of the linearized measurement equations (25):

$$\delta \begin{bmatrix} u'_{ij} \\ v'_{ij} \end{bmatrix} = J\delta \begin{bmatrix} t_i \\ \alpha_i \\ \beta_i \\ \gamma_i \\ P_j \end{bmatrix} \Rightarrow \delta \begin{bmatrix} t_i \\ \alpha_i \\ \beta_i \\ \gamma_i \\ P_j \end{bmatrix} = (J^T J)^{-1} J^T \delta \begin{bmatrix} u'_{ij} \\ v'_{ij} \end{bmatrix} \tag{25}$$

For all the unknowns to be observable, the Jacobian matrix J must be well-conditioned. Capturing a certain geometry change by tracking the helicopter simultaneously at the three cameras yields enough observability for the LS estimate. Also, to ensure proper convergence in the nonlinear LS iteration, bundle adjustment is seeded with multiple sets of initial estimates centered around the SFM solution to avoid converging to a local minimum.

5 Field Demonstration

5.1 Experimental Setup

The current prototype system consists of a helicopter platform and a ground-based camera array, Fig. 3. The camera array includes three compact digital cameras (Point Grey Firefly2 cameras, Fig. 2(a), using a Firewire interface) all connected to a single PC. An image from each camera is captured, nearly simultaneously, at a resolution of 640×480 in an 8-bit grayscale format at a rate of $20Hz$.

5.2 Tracking

The tracking algorithm based on the background differencing method was implemented in the field on each camera to track a common helicopter. Fig.

Fig. 3. Experimental setup

4(a) shows an image from one of the cameras during the test. The black box is the tracking marker centered at the estimated helicopter location and the thin white larger box is the search window of the background differencing method.

This particular flight test was conducted in an open field on Stanford's campus next to a road where moving cars and walking people constantly came in and out of the scene. While the windowed background differencing-only method frequently failed to track a low flying helicopter in such a busy environment, the Kalman filter was able to maintain the lock on the helicopter during the flight.

Fig. 4(b) shows the resulting helicopter trajectory in the image coordinate for camera 1. The solid lines show the helicopter trajectory tracked by the Kalman filter. The dashed lines show the helicopter trajectory manually post-traced in the logged images as true reference. Although the Kalman filter was able to keep tracking the helicopter, the tracking markers were sometimes lagging in tracking the helicopter when the helicopter accelerated faster than the pre-defined dynamic model in the Kalman filter equations; we believe that fine tuning the process noise covariance will further improve performance. The mean errors between the Kalman filter and the true references were roughly 7.5 pixels, as shown in Table 1(a).

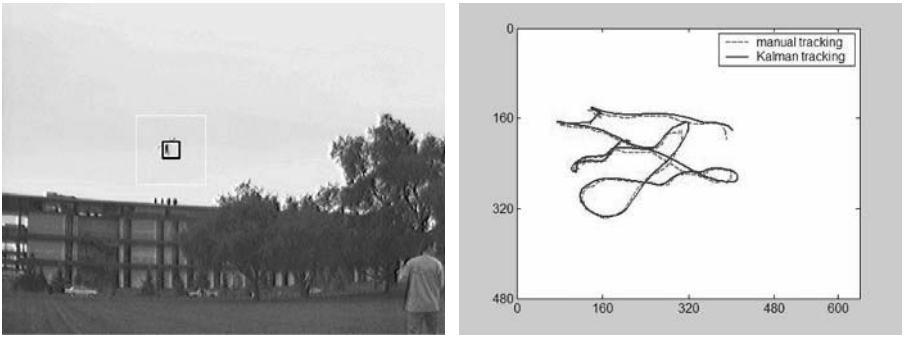
Table 1. Error tables

(a) Tracking errors in pixels

	mean(pixel)	std(pixel)
Camera 1	6.9	5.0
Camera 2	7.2	5.2
Camera 3	8.3	6.2

(b) Localization error

	x(m)	y(m)	z(m)
mean	-0.17	1.39	0.27
std	1.07	0.99	0.52



(a) View of a camera actively tracking. (b) Results of tracking.

Fig. 4. Tracking the helicopter in a busy scene.

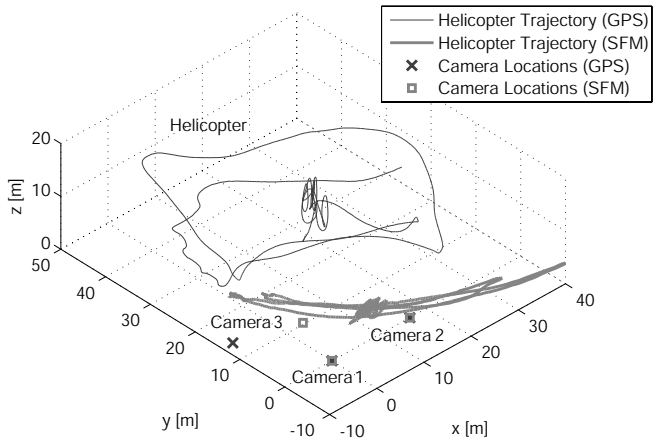
5.3 Localization

To check the validity of the localization algorithm the results from the calibration algorithm are compared with DGPS data, Fig. 5(a) and Fig. 5(b). Fig. 5(a) shows the results from SFM which is used to feed bundle adjustment. As the plot shows, care needs to be taken in picking points to initialize SFM because of the near-perspective assumption. Fig. 5(b) shows that the result from bundle adjustment follows DGPS pretty well. There are some small offsets that are probably due to the assumptions made in section 4.2. There is also some small variations in the trajectory reported by the vision system which likely result from small errors in the helicopter tracking system. Overall, the vision results match fairly well with the DGPS data. The errors are reported in Table 1(b).

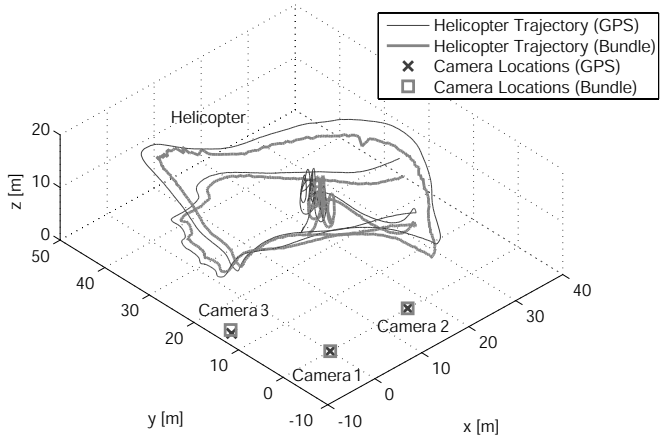
6 Conclusions

The self-surveying and tracking camera array presented in this paper produces an effective localization system that extends ideas from SFM and bundle adjustment. This is then combined with stereo tracking methods to generate a least squares measurement of the helicopter's location.

The results presented document the tracking performance of this method in a field environment for a series of cameras whose extrinsic parameters are not known *a priori*. The calibration performance was tested against DGPS and was found to have less than a $2m$ RMS tracking error. The $20Hz$ update rate of the system is comparable to DGPS, and we obtain a tracking latency of less than $100ms$. This makes it feasible to use this system as part of an autonomous flight controller.



(a) Structure from Motion.



(b) Bundle adjustment.

Fig. 5. 3-D plots of DGPS vs. ...

Because of the near-perspective assumption, it is better to run SFM on fewer points where the helicopter is near the center of the image as opposed to a large set of data where the helicopter's route spans the entire image. To make this procedure more robust, a paraperspective SFM [3] or a perspective SFM [13] can be used to initialize bundle adjustment.

Recent autonomous hover flights have demonstrated the capability of this system for real-time fielded operations. [15] Future work will test its use for acrobatic flights, find ways to maximize the flight volume, and make the system more robust to dropouts where the helicopter leaves one camera's field of view.

Acknowledgment

We would like to thank Ben Tse, our helicopter pilot, for his help and advice during our numerous experiments. We would also like thank Gary Bradski for helpful conversations about this work.

References

1. P. Maybeck, Stochastic Models, Estimation, and Control, Volume 1. Academic Press, Inc, 1979.
2. E. Trucco and A. Verri, Introductory Techniques for 3-D Computer Vision. Prentice Hall, 1998.
3. C. Poelman and T. Kanade, "A paraperspective factorization method shape and motion recovery," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 3, Mar. 1997.
4. D. A. Forsyth and J. Ponce, Computer Vision: A Modern Approach. Prentice Hall, 2003.
5. M. Whalley, M. Freed, M. Takahashi, D. Christian, A. Patterson-Hine, G. Schulein, and H. R., "The NASA / Army Autonomous Rotorcraft Project," in Proceedings of the American Helicopter Society 59th Annual Forum, Phoenix, Arizona, 2003.
6. S. Saripalli, J. Montgomery, and G. Sukhatme, "Visually-guided landing of an autonomous aerial vehicle," IEEE Transactions on Robotics and Automation, 2002.
7. J. M. Roberts, P. I. Corke, and G. Buskey, "Low-cost flight control system for a small autonomous helicopter," in IEEE International Conference on Robotics and Automation, 2003.
8. P. Liang, P. Chang, and S. Hackwood, "Adaptive self-calibration of vision-based robot systems," IEEE Transactions on Systems, Man and Cybernetics, vol. 19, no. 4, pp. 811–824, July 1989.
9. G. Mayer, H. Utz, and G. Kraetzschmar, "Towards autonomous vision self-calibration for soccer robots," Proc. of the Intelligent Robot and Systems (IROS) Conference, vol. 1, pp. 214–219, 2002.
10. J. Knight and I. Reid, "Self-calibration of a stereo rig in a planar scene by data combination," In Proc. of the Internatoinal Conference on Pattern Recognition, pp. 1411–1414, Sept. 2000.
11. Z. Zhang, "Motion and structure of four points from one motion of a stereo rig with unknown extrinisic parameters," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 12, Dec. 1995.
12. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," IEEE Aerospace and Electronic Systems Magazine, vol. 19, no. 1, Jan. 2004.
13. M. Han and T. Kanade, "Perspective factorization methods for euclidean reconstruction," Carnegie Mellon, Tech. Rep. CMU-RI-TR-99-22, Aug. 1999.
14. S. Thrun, G. Bradski, and D. Russakoff, "Struction from motion," Feb. 2004, lecture Notes from CS223b.
15. A. Ng, A. Coates, M. Diel, V. Ganapathi, J. Schulte, B. Tse, E. Berger, and E. Liang, "Inverted autonomous helicopter flight via reinforcement learning," International Symposium on Experimental Robotics, 2004.