1st International Conference on Sensing Technology
November 21-23, 2005 Palmerston North, New Zealand

A Micro Aerial Vehicle Motion Capture System

Christel-loic Tisse, Thomas Fauvel and Hugh Durrant-Whyte
ARC Centre of Excellence for Autonomous Systems
The University of Sydney NSW 2006, Australia
c.tisse@cas.edu.au

Abstract

In this paper, we report on the design of a 6 Degrees-Of-Freedom (6-DOF) motion tracking system for a Micro-Aerial-Vehicle (MAV) using stereo-vision. The pose measurement is performed by tracking the 3D positions of a marker (three targets) attached to the MAV. This task involves determination of optimal sensor placement for the stereo system and an orientation calculation algorithm. Experiments were conducted to validate the set-up against defined accuracy requirements.

Keywords: unmanned aerial vehicle, 3D motion tracking, pose estimation, stereo-vision, sensor placement.

1 Introduction

Pose estimation is of critical importance in autonomous robotic Micro-Air-Vehicle (MAV) research as it is the principal measurement used in localization of collected data, machine control and dynamic modeling. The approach in this paper involves using two cameras located optimally in a room to track the orientation and position of an indoor micro-helicopter, such as *Mosquito* (Fig.1). The purpose is to extract accurately its position and orientation in response to control inputs. This flight data will eventually lead to a dynamic model of the platform, using system identification.

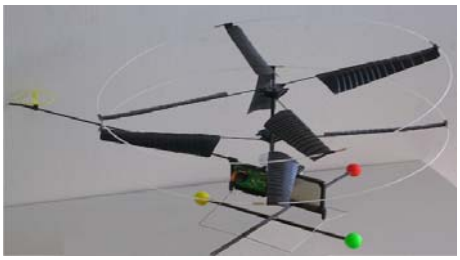

Figure 1: *Mosquito*, our MAV research platform: Rotor Ø = 30*cm* , Weight = 55*g*, and Payload ≈ 5*g*.

1.1 Pose sensing requirements

The required accuracy of the motion capture system is defined by *Mosquito*'s flight abilities. Unlike regular micro-helicopter, *Mosquito* is passively stable in hover and during forward flight, which restricts the range of orientation. Roll and pitch angles are typically within +/-20º. Yaw angle is free. With regards to the position accuracy, we estimated 2% of the vehicle size (i.e. 10*mm* along the 3 axes) to be suitable for dynamics modeling. Finally, the video frame-rate needed to track 3D movement depends on the *Mosquito*'s velocity. Its maximum translatory velocity is $1.5 m.s^{-1}$ horizontally and $2.5 m.s^{-1}$ vertically (*i.e.* free fall), and its angular velocity is $90 deg.s^{-1}$. Preliminary flight test data suggests also that a minimum frame-rate of $15 frames.s^{-1}$ would need to be adopted if *Mosquito* was visually controlled.

1.2 Background on pose trackers

In the literature, there have been several techniques proposed for tracking the pose of moving objects. See [1,2,3] for a more in-depth review of existing technologies. Exotic solutions include: (i) ultrasonic systems designed for short range distances; (ii) RF beacons (inaccurate and battery-powered); and (iii) laser scanner (very expensive and too bulky for *Mosquito*). The more popular approaches are categorized in the following sub-sections.

1.2.1 Inertial solution

Miniaturized inertial sensors have been used to develop tracking systems. They provide relatively high-bandwidth motion measurement with very low noise. They measure directly the motion derivatives and require integration over time, which will cause the orientation and position to gradually drift. Even with very light MEMS inertial sensors, drift-corrected hybrid tracking technology remains unsuitable for the payload of our MAV (≈5*g*).

1.2.2 Magnetic solution

After decades of evolution, magnetic tracking is a very mature technology, and has been used in many applications (*e.g.* mini rotorcraft stabilization [4]). They have several advantages: 6-DOF tracking, small sized sensor and no light-of-sight requirement between source and sensor. However, the very short range of magnetic field (typically less than 70cm), the high sensitivity to metallic environment (despite improved calibration [5,6]), and the existing wired solution are very much an issue for micro flyer.

1.2.3 Optical solution

Various optical systems have been designed. Three approaches can be distinguished: (i) the *outside-in*, where cameras look at targets on the frame of interest, (ii) the *inside-out*, where the camera is on-board, and (iii) the *hybrid* based on a pair of ground and on-board cameras. The *hybrid*, introduced recently [7], requires six coloured markings (five are placed on the moving object and one is located on the ground camera) for accurate pose estimation. Development of lightweight, onboard solution is under active research and remains a significant challenge in micro aerial vehicle. *Outside-in* represents the least invasive and potentially the most precise solution, though most commercial systems require a large, complex and expensive set-up. For all these reasons, we considered to build our own *outside-in* motion capture system using stereo vision.

1.3 Contribution and paper outline

This paper examines the application of stereo vision techniques to the 6-DOF motion tracking of *Mosquito*. The proposed system consists of only two cameras and a passive marker embedded on *Mosquito* (*i.e.* three colored targets spaced equally a few centimeters apart). The rest of the paper is organized as follows. Section-2 defines the optimal sensor placements. Section-3 reviews the triangulation approach and describes the equations for the derivation of the attitude angles. We then present in Section-4 some experimental results that demonstrate the performance of our MAV motion capture system. The paper concludes with a discussion of future work.

2 Optimal sensor placement

The problem of optimal sensor placement for single and multiple vision sensors network has been investigated [8,9,10,11]. Close-range *photo-grammetry* (or *videogrammetry*) – technique of measuring 3D objects from pictures – has also studied a number of issues relating to reconstruction optimization by carefully positioning the vision sensors [12,13,14]. We present here a pragmatic approach to the problem of stereo vision sensor design, based on the minimization of the triangulation error.

2.1 Focusing on the geometric aspect

C.Cowan *et al.* [8] have identified the following criteria for acceptable viewed points that affect sensor placement: (i) resolution, (ii) focus, (iii) field-of-view (FOV), (iv) visibility, and (v) view angle. The optimum solution can be met by using small spherical targets. Resolution and view angle summarize into the accuracy of localizing a circular shape in the image. An explicit set-up allows the derivation of an upper bound to this accuracy (*cf.* error in 3D point position). This aspect is addressed later in this section. In our application, FOV and focus restrict the study volume, but not the possible camera positions. With regard to visibility, the targets' positions on the micro-helicopter are chosen to minimize occlusion in the cameras' FOV. *Mosquito*'s body (made of thin carbon tubes) is indeed expected to partially occlude the targets. Setting the viewing angle (of both cameras) incident to the most frequent plane drawn by the three targets resolves the problem of inter-occlusion.

The determination of 3D positional accuracy from a detected circle in stereo vision images has been discussed in [12,15,16,17,18]. The geometrical arrangement of the sensors is also very important for minimizing the localization error [18]. This error can be expressed as an uncertainty in detecting the spherical target boundary from each camera. In 3D space, this uncertainty is represented by the intersection of two cones whose main axes intersect in the correct 3D point position.

2.2 Triangulation error and position of the cameras

According to [17], the uncertainty of reconstruction increases as the angle between the rays of each camera (*i.e.* convergence angle) becomes narrower. Chiou *et al.* [16] derived the optimal camera geometry of a trinocular vision system by analyzing a non-ambiguity probability of potential feature correspondences. Error analysis of stereo measurements has also been conducted from the study of stereoscopic system geometry and the characteristics of the image [15]. We exploit here similar approach to verify which convergence angle leads to a minimum triangulation error.

Assuming two pinhole cameras (with respective centre of projection A and B) and a 3D point P, the maximum triangulation error can be equally treated in the *epipolar* plane ABP as illustrated in Fig.2.

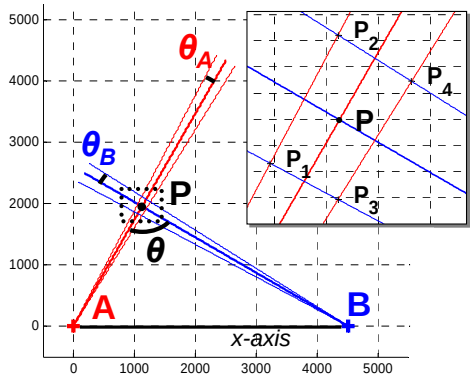

Figure 2: Triangulation error: the region delimitated by the points P_1, P_2, P_3, and P_4 illustrates the shape of the uncertainty region. The position error of the reconstructed points comes from the intersection of error cones. The maximum error lies in the plane *(ABP)*.

We define the triangulation error ε_p as the maximal distance between the point P and the points $P_{n=1,2...4}$. A *generate-and-test* method [11] is then used to compute $min(\varepsilon_p)$ as a function of the convergence angle θ, the distance $\overline{AB}$ between the cameras, and different error cones angles θ_A and θ_B. Fig.3 shows some simulation results that we obtained for $\overline{AB}$=4500mm and θ_A=θ_B=π/1000 (angles are expressed in *radians*, unless explicitly specified). The error ε_p is minimum for every sensor placement that features a convergence angle θ=π/2. This is represented in Fig.3 as a half-circle centered between A and B.

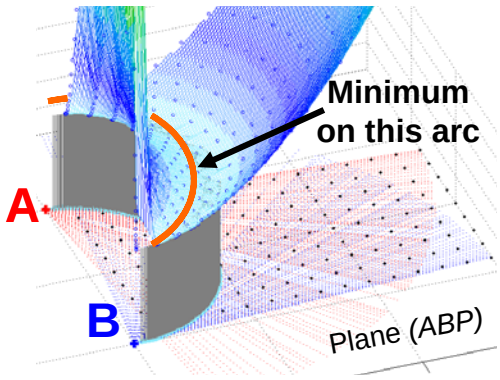

Figure 3: Estimation of the triangulation error. Points P are sampled on a 5×5mm grid in the plane *(ABP)*; the angle θ varies within the range]0,π[. A minimum triangulation error of +/-8mm is reached for θ_0=π/2, independently of variations in $\overline{AB}$,θ_A, and θ_B.

2.3 Motion detection maximization and stereoscopic system orientation

We have demonstrated the optimal relative positions of the cameras. Another critical parameter is the global orientation of the stereovision system. In particular, we want to determine the sensor placement that maximizes the projections of the target translation motion vector (T=[T_x, T_y, T_z]^T) in both camera planes.

We assume that the two cameras are pointing towards the center of a world reference frame (O,x,y,z). Let ψ and ϕ be respectively the azimuth angle and the polar angle representing the optical axis orientation (represented by the unit vectors $u_A(\phi_A,\psi_A)$ and $u_B(\phi_B,\psi_B)$) of the cameras A and B. To estimate the optimal system orientation, we compute the 2D projections (T_A and T_B) of the translation vector T in the image plane of the two cameras. The possible set of azimuth and polar angles must verify the condition implied by the singular angle of convergence, θ_0=π/2. This condition is satisfied for

$$\sin(\phi_A)\sin(\phi_B)\cos(\psi_A-\psi_B)+\cos(\phi_A)\cos(\phi_B)=0. \qquad (1)$$

In general, there is no guarantee that Eq.1 yields a solution as detailed in Table-1.

Table 1: Possible solutions to equation (1).

	$\phi_B = 0$	$0 < \phi_B < \pi/2$	$\phi_B = \pi/2$
$\phi_A = 0$	No solution	No solution	ψ_A, ψ_B free
$0<\phi_A<\pi/2$	No solution	$\psi_B = \psi_A - \text{acos}\left(\frac{-1}{\tan\phi_A \tan\phi_B}\right)$ only for $\phi_A + \phi_B \in [\pi/2,\pi[$	$\psi_B = \psi_A \pm \pi/2$°
$\phi_A = \pi/2$	ψ_A, ψ_B free	$\psi_B = \psi_A \pm \pi/2$	$\psi_B = \psi_A \pm \pi/2$

Let us call R_A and R_B the rotation matrices from the world reference frame to the image planes. Thus we can write

$$\begin{cases} T_A = KR_AT \\ T_B = KR_BT \end{cases}, \text{ with } K = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right]. \qquad (2)$$

For a small displacement ΔT_x, ΔT_y, or ΔT_z, of the target along respectively the x, y or z directions, the magnitude of the projected vector T_A (and similarly for T_B) has the form

$$\begin{cases} |T_A|_{(\Delta Tx)} = \delta\sqrt{\left(1-(\sin(\phi_A)\cos(\psi_A))^2\right)} & if\ \Delta Tx \\ |T_A|_{(\Delta Ty)} = \delta\sqrt{\left(1-(\sin(\phi_A)\sin(\psi_A))^2\right)} & if\ \Delta Ty \\ |T_A|_{(\Delta Tz)} = \delta\sqrt{\left(1-(\cos(\psi_A))^2\right)} & if\ \Delta Tz \end{cases}. \qquad (3)$$

As we do not want to favor one particular direction for the detection of the target displacement (isotropic optimization), the best sensor placement corresponds to the configuration $\{\phi_A,\phi_B,\psi_A,\psi_B\}_0$ that maximizes the function "$|T_A|_{(\Delta Tx)}$+ $|T_B|_{(\Delta Tx)}$+ $|T_A|_{(\Delta Ty)}$+ $|T_B|_{(\Delta Ty)}$+ $|T_A|_{(\Delta Tz)}$+ $|T_B|_{(\Delta Tz)}$".

Simulations using equations (1), (2), and (3), show that the optimal orientation of the stereoscopic system occurs at $\{\phi_A$=ϕ_B=ψ_A=$\pi/4;\psi_B$=-3$\pi/4\}$. Locations of the sampled orientations and the optimal sensors placement are illustrated in Fig.4.

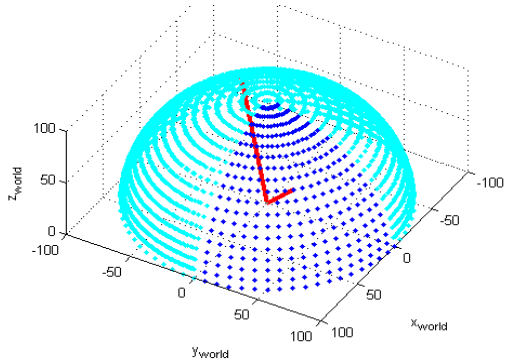

Figure 4: Estimation of the optimal stereo vision system orientation. The dots indicate the location of the sampled orientations. Simulations were performed for ϕ_A and $\psi_A \in [0,\pi/2]$. The solid lines represent the optical axes of the cameras in their optimal positions.

3 Motion capture approach

The pose estimation can be defined as: finding the orientation $r=(\varphi,\theta,\psi)$ (φ=roll, θ=pitch, and ψ=yaw) and relative position $P(x_0,y_0,z_0)$ of the moving marker (attached to our indoor micro-helicopter) with respect the world reference frame, as well as translational and angular velocities in "real-time". Using color detection and simple triangulation [19,20], the 3D trajectory of the marker can be recovered accurately with respect to the calibrated stereo-vision system (located on the ground). Calibration of the cameras is necessary to determine extrinsic parameters such as the rotation and translation relationship between each camera. Estimation of the full 6-DOF pose of *Mosquito* requires the tracking of at least three targets (*i.e.* colored polystyrene balls). Though the use of several points for pose estimation is an established technique in augmented reality (*cf.* cinema post-production, where virtual objects must be added to real-world footage), it has never been tested for tracking online the 3D orientation of a MAV.

3.1 Orientation computation from 3D points position

Let us consider a marker composed of three points (*1*), (*2*), and (*3*), in the world reference frame, as illustrated in Fig.5. Its orientation $r=(\varphi,\theta,\psi)$ can be derived as follows.

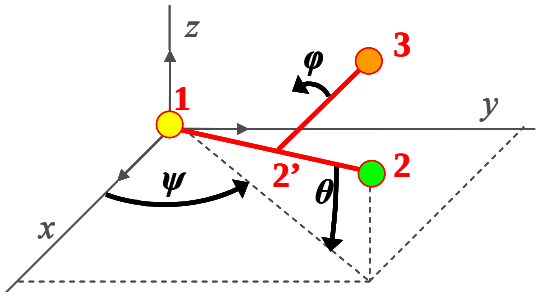

Figure 5: The orientation $r=(\varphi,\theta,\psi)$ of the rigid marker. r is defined by the 3D position of the three spherical targets in the world reference frame. φ and $\theta \in [-\pi/2,\pi/2]$, and $\psi \in [0,2\pi]$.

Let the point (*2'*) be the orthogonal projection of the point (*3*) on the line segment (*12*). Let R be the rotation matrix from the body-fixed frame of the marker to the world reference frame. Let (x_j,y_j,z_j) define the Cartesian coordinates of a point (j) in the world reference frame. We can express the following roll φ, pitch θ, and yaw ψ angles:

$$\sin\varphi = \left(\frac{(z_3 - z_{2'})}{\overline{2'3}} \frac{\overline{12'}}{\sqrt{(x_{2'} - x_1)^2 + (y_{2'} - y_1)^2}} \right) \text{ for } \theta \neq \pm\frac{\pi}{2} \quad (4)$$

$$\tan\theta = \left(\frac{-(z_2 - z_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} \right), \ \theta \neq \pm\frac{\pi}{2} \quad (5)$$

$$\begin{cases} \psi = \text{sign}(y_2 - y_1) \times \left(\frac{\pi}{2}\right) \ if \ x_2 = x_1 \\ \psi = \text{atan}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \ \ if \ x_2 > x_1 \\ \psi = -\pi + \text{atan}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) else \end{cases} \quad (6)$$

3.2 Orientation accuracy versus triangulation error

The accuracy in the measurement of the marker's orientation is directly related to the triangulation error in the estimation of the 3D points' position. Assuming a triangulation error $\varepsilon \approx$ +/-8*mm* and an inter-distance of 160*mm* between the spherical targets of the marker, we assessed the error on the attitude estimation using similar *generate-and-test* approach adopted in Section-2.2. Roll and pitch angles vary between +/-20º and yaw angle within the full range (5º stepwise). We observed a maximum orientation error of ~4.45° for roll, ~5.60° for pitch, and ~5.30° for yaw. We also noticed that these errors increase as the range of roll/pitch angles increases.

4 Experiments

A stereo vision system has been prototyped to validate the proposed design concept of MAV motion capture system. Flight motion simulation was conducted in our laboratory to measure its accuracy in a real indoor environment (see Fig.6).

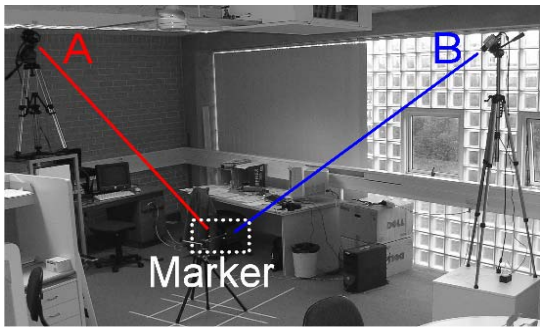

Figure 6: Experimental setup (two cameras and the marker).

4.1 Experimental setup

The marker is made of three colored polystyrene spheres (Ø≈15*mm*) equally spaced by 160*mm* (Fig.1), and weights 2*g*. Each sphere is painted with a different bright color (yellow, green, or orange) to facilitate the blob detection process. In addition, the stereo vision system includes two high-quality digital cameras to capture sharp (*i.e.* free of color artifacts) color images: a *Hitachi* 3-CCD HVD30 camera (3-CCD technology), and a *HanVision* HVDUO-10M camera (*FoveonX3* technology). The images are synchronically recorded at a resolution of 640×480 pixels in an 8-bit RGB format at a rate of 15*frames/s*.

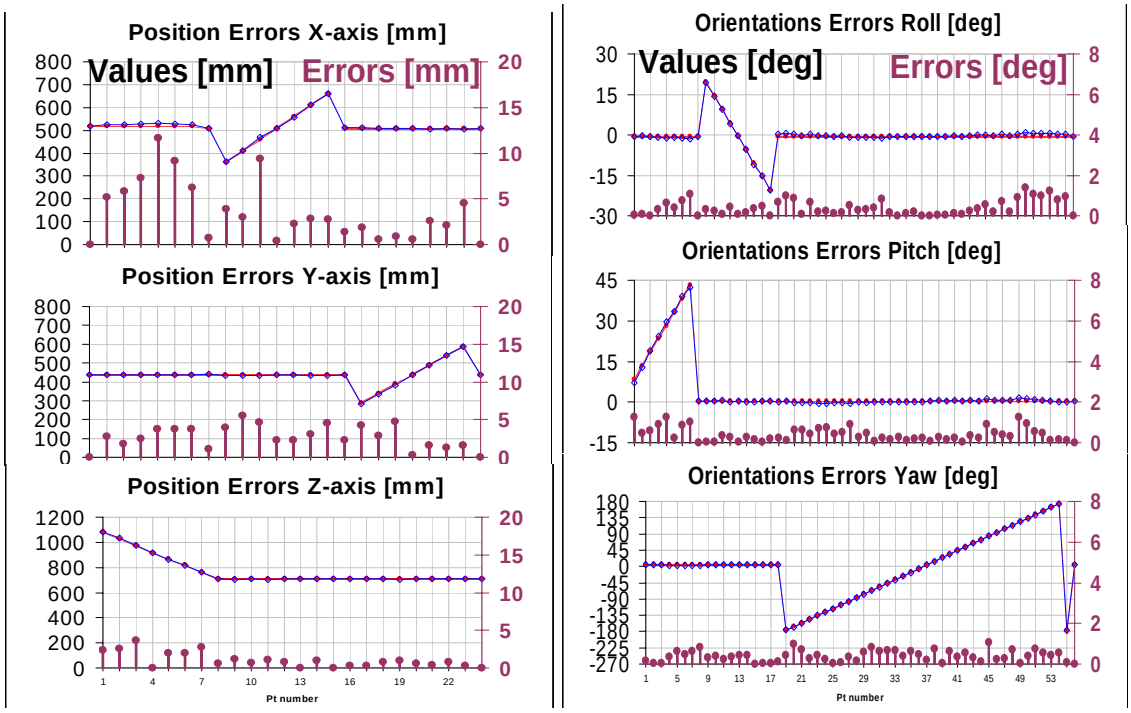

Figure 7: Results of the 3DOF position and 3DOF orientation computation with the stereoscopy-based pose estimation method.

Each camera was connected to a separate computer. The ability to locate and track the marker in "real-time" is limited by the computers (Pentium 4, 1.6GHz machines) and frame-grabbers (*Matrox* Meteor-II and *ActiveSilicon* Phoenix-LVDS). The FOV of both cameras is approximately 30°. The cameras are positioned according to the optimal sensor placement that we demonstrated in Section-2. A 4.5*m* baseline between the two cameras enables a tracking volume of $2\times1\times1.5m^3$. The calibration process was performed using a $1.5\times1.5m^2$ chessboard and the online *Matlab* toolbox available from Caltech Institute [21].

4.2 Pose measurement accuracy

To assess the pose accuracy of our stereoscopy-based pose tracker, we considered simple 3D flight trajectories of *Mosquito* in the viewing field of the cameras. The sequences of 3D motion signals were manually recorded at randomly chosen locations. In each sequence, the marker attached to *Mosquito* undergoes six single motions: three translations of 50*mm/frame* along *x*, *y*, and *z* axes, followed by two rotations of 5*°/frame* about *x* and *y* axes, and one rotation of 10*°/frame* about *z* axis. Fig.7 shows, as example, the results of pose measurement accuracy we obtained in the centre of the motion capture *arena*. The average absolute error in estimating the 3DOF position is in the order of 4.57*mm* (or lower depending on the motion direction). The standard deviation in the position error measurements is 3.73*mm*. The maximum observed positional error was 11.60*mm*. The average error in measuring the 3DOF orientation is about 0.39° for roll, 0.48° for pitch, and 0.40° for yaw. The reported standard deviation in the orientation measurements is relatively high (σ_{roll}=0.36°, σ_{pitch}=0.56°, σ_{yaw} =0.26°), despite the large number of measures. A deeper analysis of the whole collected database shows that inaccurate pose estimates are mostly due to imprecision in our direct colored blob detection technique, sensitive to partial occlusion and shadow.

5 Concluding remarks and future work

Exploring control methodologies and modeling the dynamics of micro unmanned air vehicles require new technologies for studying in-flight motion. The stringent constraints placed on MAV (*e.g.* payload capability of a few grams) make existing motion tracking system impractical.

In this paper, we have reported the development of a stereoscopic optical pose tracker for *Mosquito*, our robotic micro flyer. The stereo system is based on two intrinsically calibrated cameras and the tracking of three lightweight, colored, spherical targets (attached to *Mosquito*). The placement of the stereo vision cameras was optimally determined through simulation. Initial experiments on a tethered platform have shown that the 3DOF position/orientation measurements delivered by the proposed pose estimation method are accurate enough to be used for motion recovery of *Mosquito*. The typical absolute

pose error is about 4.6mm in position, and less than 0.5 degree in orientation.

In future work, we intend to improve the performance of the pose tracker by adding robust *Kalman* filtering method and a more accurate target (circle) detection algorithm. We expect then to validate this approach for *Mosquito*'s flight dynamics modeling.

6 Acknowledgement

This work is supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales state government.

7 References

[1] Borenstein, J., Everett, H.R. and Liqiang, F., *Where am I? -- Sensors methods for mobile robot positioning*, University of Michigan, Ann Arbor, MI (1996).

[2] G. Welch and E. Foxlin, "Motion Tracking: No silver bullet, but a repectable arsenal," *IEEE Computer Graphics and Application, Special issue on Tracking,* 22(6), pp. 24-38 (2002).

[3] C. Youngblut, R.E. Johnston, S.H. Nash, R.A. Wienclaw and C.A. Will, *Review of Vitual Environment Interface Technology*, Chapter 3, Institue for Defense Alalyses (IDA), Alexandria, VA (1996).

[4] P. Castillo, A. Dzul and R. Lozano, "Real-Time Stabilization and Tracking of a Four-Rotor Mini Rotorcraft," *IEEE Transactions on Control Systems Technology,* 12(4), pp. 510-516 (2004).

[5] M. Ikits, J.D. Brederson, C.D. Hansen and J.M. Hollerbach, "An improved calibration framework for electromagnetic tracking devices," In *Proceedings of the IEEE Annual International Symposium on Virtual Reality,* Yokohama, Japan, pp. 63-70 (2001).

[6] G. Zachmann, "Distortion Correction of Magnetic Fields for Position Tracking," In *Proceedings of Computer Graphics International*, pp. 213-220 (1997).

[7] E. Altug, J.P. Ostrowski and C.J. Taylor, "Control of a Quadrotor Helicopter Using Dual Camera Visual Feedback," *The International Journal of Robotics Research,* 24(5), pp. 329-341 (2005).

[8] C.K. Cowan and P.D. Koveski, "Automatic Sensor Placement from Vision Task Requirements," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 10(3), pp. 407-416 (1988).

[9] S.O. Mason and A. Grun, "Automatic Sensor Placement for Accurate Dimensional Inspection," *Computer Vision and Image Understanding,* 61(3), pp. 454-467 (1995).

[10] G. Olague and R. Mohr, "Optimal camera placement for accurate reconstruction," *Pattern Recognition,* 35(4), pp. 927-944 (2002).

[11] K.A. Tarabanis, P.K. Allen and R.Y. Tsai, "A survey of sensor planning in computer vision," *IEEE Transactions on Robotics and Automation,* 11(1), pp. 86-104 (1995).

[12] C. Fraser, "Optimization of precision in close range photogrammetry," *Photogrammetric Engineering and Remote Sensing,* 48(4), pp. 561-570 (1982).

[13] C. Fraser, "Automated Vision Metrology: A Mature Technology for Industrial Inspection and Engineering Surveys," In *6th South East Asian Surveyors Congress,* Fremantle, Western Australia, (1999).

[14] G. Ganci and H. Handley, "Automation in Videogrammetry," *International Archives of Photogrammetry and Remote Sensing,* 32(5), pp. 53-58 (1998).

[15] J. Amat, M. Frigola and A. Casals, "Selection of the best stereo pair in a multi-camera configuration," In *Proceedings of the IEEE International Conference on Robotics and Automation,* Wahington, DC, vol. 4, pp. 3342-3346 (2002).

[16] R.M. Chiou, C.H. Chen, K.C. Hung and J.Y. Lee, "The optimal camera geometry and performance analysis of a trinocular vision system," *IEEE Transactions on Systems, Man and Cybernetics,* 25(8), pp. 1207-1220 (1995).

[17] Hartley, R. and Zisserman, A. *Multiple View Geometry in computer vision* (2nd ed.), Cambridge University Press, Cambridge, UK (2003).

[18] J.H. Lee, T. Akiyama and H. Hashimoto, "Study on Optimal Camera Arrangement for Positioning people in Intelligent Space," In *IEER/RSJ International Conference on Intelligent Robots and System,* Lausanne, Switzerland, vol. 1, pp. 220-225 (2002).

[19] Faugeras, O., *Three-dimensional computer vision: a geometric viewpoint,* MIT Press, Cambridge, MA (1993).

[20] I. Mitsugami, N. Ukita and M. Kidode, "Estimation of 3D gazed position using view lines," In *Proceedings of the 12th International Conference on Image Analysis and Processing,* Mantova, Italy, pp. 466-471 (2003).

[21] J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab", visited on May 2005, www.vision.caltech.edu/bouguetj/calib_doc/.