

# A vision-based method for autonomous landing of a rotor-craft unmanned aerial vehicle

doi:10.1533/abbi.2006.0026

Z. Yuan, Z. Gong, J. Chen and J. Wu

*School of Mechatronics Engineering and Automation, Shanghai University, 149 Yanchang Road, Shanghai, People's Republic of China*

**Abstract:** This article introduces a real-time vision-based method for guided autonomous landing of a rotor-craft unmanned aerial vehicle. In the process of designing the pattern of landing target, we have fully considered how to make this easier for simplified identification and calibration. A linear algorithm was also applied using a three-dimensional structure estimation in real time. In addition, multiple-view vision technology is utilized to calibrate intrinsic parameters of camera online, so calibration prior to flight is unnecessary and the focus of camera can be changed freely in flight, thus upgrading the flexibility and practicality of the method.

**Key words:** Autonomous landing, identification, homography, multiple view vision, calibration.

## INTRODUCTION

Unmanned aerial vehicles (UAV) have become very important in modern times in such fields as military reconnaissance, geography mapping, traffic management, aerial safety monitoring, air live broadcasting, environmental protection, urban planning, remote measurement and remote sensing, and surveillance of natural disasters (Saripalli and Montgomery 2002; Gong 2003; Saripalli *et al.* 2003). Rotor-craft UAV is an important type of UAV that is capable of vertical takeoff and landing and can stay in air for a fixed-point observation job, thus boasting sound maneuverability. For rotor-craft UAVs to function successfully, autonomous landing is a crucial capability. This is particularly difficult because of the inherent instability of rotor-craft UAVs near the ground (Shakernia *et al.* 1999), and a flight control system is needed to promptly adjust the control parameter.

At present, both global positioning system (GPS) and inertial measurement unit (IMU) are used to make up the flight control and navigational system for autonomous flight of most rotor-craft UAVs, which can satisfy control requirements for a long-range cruise along a set track and can provide information on the probable location of its

landing point. As the major sensor for the main task undertaken by a rotor-craft UAV, this vision sensor can achieve such tasks as monitoring and shooting in air, and can also provide a natural sensing modality for object detection and landing. Vision-based autonomous landing technology has recently become an active topic of research (Shakernia *et al.* 1999; Sharp and Shakernia 2001; Shakernia and Vidal 2002; Saripalli and Montgomery 2002; Saripalli *et al.* 2003). Shakernia and Vidal (2002) have worked out the UAV motion estimation method of multiple views that makes use of the rank deficiency of the multiple-view matrix, and Sharp and Shakernia (2001) have also worked out a motion estimation method based on a dual view. Still, for Saripalli, it is an estimation method to enable a helicopter to land on moving targets that is based on vision technology in which moments of inertia are applied to identify targets. The Kalman filter is used in target tracking, and a track controller is employed to enable the helicopter to land on a target (Saripalli and Sukhatme 2003).

Our work has been inspired by the work mentioned above, with our focus being on upgrading the flexibility and practicality of the vision method. We propose a unique and simple idea for designing and identifying a landing target, thus obtaining a frame rate of 25 Hz. Another improvement is that calibration can be done online. In practice, we need to change the focus of the camera during the searching and tracking task. At the same time, in order to pose estimation, the value of focus must be determined. It is a problem for most cameras to return to the exact value of focus after being changed. Multiple-view vision is implemented to solve this problem. So there is no necessity for the camera to be calibrated prior to takeoff, and its

---

Corresponding Author:

Z. Yuan

School of Mechatronics Engineering and Automation

Shanghai University, 149 Yanchang Road

Shanghai, People's Republic of China

Tel: 86-21-56334137; Fax: 86-21-56331492

Email: zpyuan@mail.shu.edu.cn

---

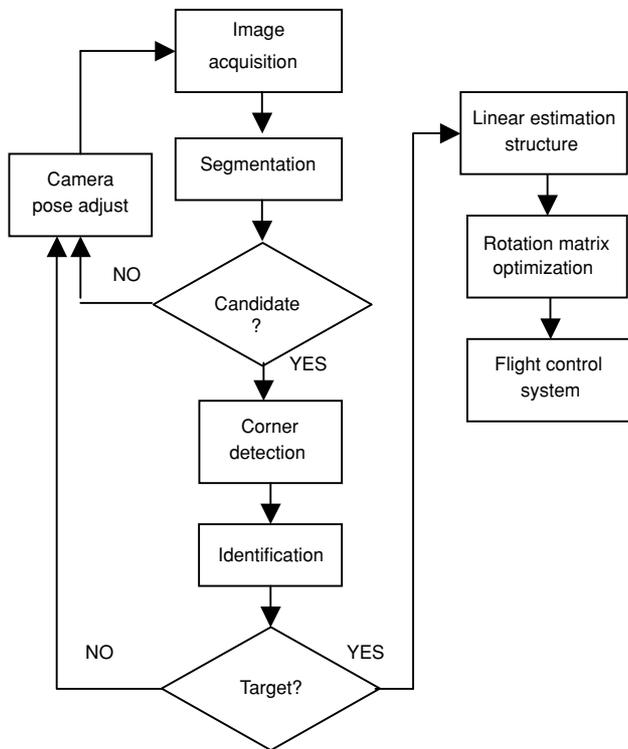


Figure 1 Pose estimation algorithm in autonomous landing.

focus can be changed in flight. Here, how to design the landing pattern is to be taken into consideration so as to provide enough corners for calibration. The UAV used in our work is a mini rotor-craft (SHU-XY-UAV 2) developed by Shanghai University, which is equipped with GPS, gradienter, magnetic compass, angular-speed top, and a two-degree-of-freedom (pitch and yaw) cradle head whose rotating angular of the pitch and yaw can be obtained through a photoelectric encoder. The article focuses on research of the vision system, while the flight control system is ignored.

**AUTONOMOUS LANDING STRATEGY**

The designed autonomous landing of our mini rotor-craft UAV is divided into three stages.

*Stage 1:* To get back to the place near the landing point through the GPS navigational system, then hover above the landing location at a height of 5–8 meters and search for a landing target. The focus of the camera would be fixed during landing.

*Stage 2:* To select five images selected from the sequence of images in which the target is included, with a distinct angle difference. The angle difference is determined using information from a gradienter and a magnetic compass. Then intrinsic calibration of the camera will be done.

*Stage 3:* To identify the target and estimate pose, then land gradually according to the guide on the pose information

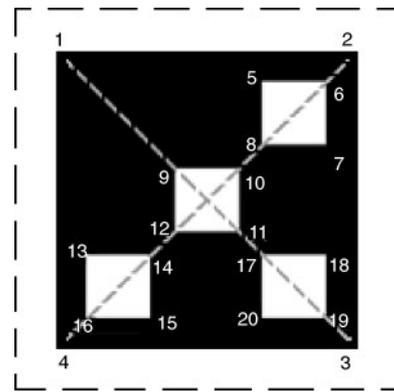


Figure 2 Pattern of target and order of corners.

from the vision system. The workflow of this vision method is shown in Figure 1.

**DESIGNING AND IDENTIFYING THE TARGET**

**Designing target**

Pattern design of the target is done to make it easier to identify and obtain image coordinates of the corners, providing enough corners for calibration is also considered. In Figure 2, there is a black square area in the white background in which four white squares are distributed in the shape of a triangle; black–white comparison is helpful for segmentation and corner extraction. Here, corner 1 is the origin of world coordinate. The direction from corner 1 to corner 4 is the positive direction of the X axis, and to corner 2 is the positive direction of the Y axis. The world coordinates of 20 corners in Figure 1 is to be measured in advance for use in pose estimation. Usually, the process for obtaining the image coordinate of the corners includes threshold segmentation, labeling, corner detecting, and identification (Sharp and Shakernia 2001; Saripalli *et al.* 2003; Rekimoto 1998). In order to upgrade the real-time effect of the method, a unique and simplified algorithm is employed in identification as illustrated in the article.

**Segmentation**

In practical application, robustness of the algorithm is a very important factor. To test the effectiveness of the method, an image taken in strong light and shadow is to be selected (see Figure 4). Since the target is designed in black and white, it is clear in the histogram (see Figure 3) that the target (see Figure 4) can be segmented by means of a global threshold. Moreover, to account for the unstable light and the changes in image brightness because the pictures were taken from different angles, a disturbing area is included in the segmentation result (see Figure 5). After threshold segmentation, the labeling is done. The labeling area is to be eliminated if its pixel number is less than the preset lower limit or more than the preset upper limit, and then the image coordinate of the pixel including  $max-x$ ,  $max-y$ ,  $mini-x$  or  $mini-y$  in the reserved candidate

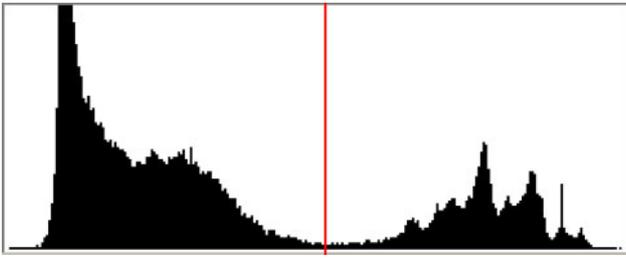


Figure 3 Histogram of grey image.

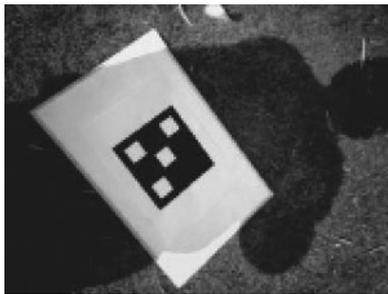


Figure 4 Sample grey image.

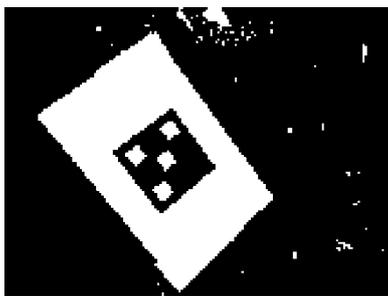


Figure 5 After threshold segmentation.

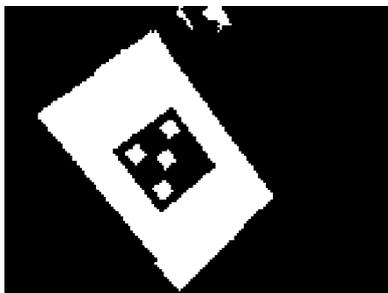


Figure 6 After filter by pixel number.

labeling area (see Figure 6) is to be recorded. According to the extreme image coordinate of the labeling area, the width/length ratio of wrap rectangular at the candidate labeling area is computed. The labeling area, whose ratio is between the lower limit and the upper limit, is to be reserved and otherwise to be eliminated (see Figure 7). After the above-mentioned processing, the reserved labeling area is called candidate areas. If there is no candidate area, the next frame is processed.

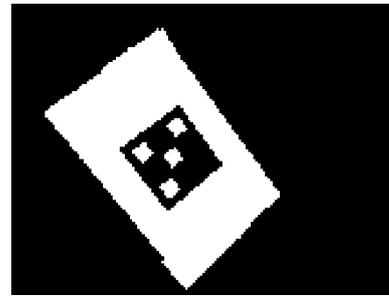


Figure 7 After filter by width/length ratio.

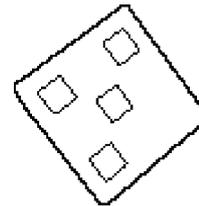


Figure 8 Edge of pattern.

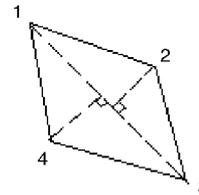


Figure 9 Case 2: extreme vertices.

### Identification

For pose estimation, corners related to the 20 corners in the segmented image should be found out. To be exact, we will make out image coordinates of the 20 corners in Figure 1.

*Step 1:* Identify four vertices of the black square pattern. First, the edge is to be extracted (see Figure 8), and then any points must be identified, including the max- $x$ , max- $y$ , mini- $x$  and mini- $y$  of the edge. Then it must be decided whether the points are the same. If they are two different points (see Figure 9), they must be the two diagonal vertices, and the other two can be realized by searching for the farthest point from the diagonal on the edge. If they are three different corners (see Figure 10), any of the two are connected to make a line; if all pixel values in eight adjacent areas at the line center are 1 (white), then the two points are diagonal vertices. The search method for other vertices is the same as those used in the case of two different points. If there are four different corners, they are taken as just four vertices (see Figure 11). Vertices of the four white squares inside can also be obtained with the same method.

*Step 2:* Identify the orders of the four vertices. In Figure 2, given centroid of the black square is 0, the pixel value of line 01's center should be 0 (black). In practice, we will decide whether all pixel values in eight adjacent areas

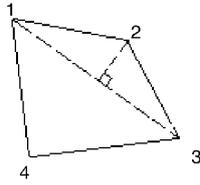


Figure 10 Case 3: extreme vertexes.

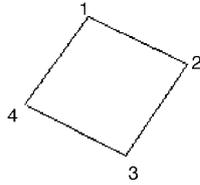


Figure 11 Case 4: extreme vertexes.

are 0; Once we get the point 1, its adjacent vertex in the counterclockwise direction should be 2, the diagonal vertex 3 and the last vertex 4. After finding points 1, 2, 3 and 4, it will be easy for us to determine the other vertexes in the rest of the four white squares.

If we cannot find 20 corners in any candidate according to this method, the candidate will not be considered as a target.

### CALIBRATION BASED ON MULTIPLE VIEWS

Before estimating the Euclidean structure, the intrinsic matrix  $\mathbf{K}$  of the camera must be known, and general model of  $\mathbf{K}$  is shown in Equation (1).

$$K = \begin{bmatrix} f_x & \gamma & u_0 \\ & f_y & v_0 \\ & & 1 \end{bmatrix} \quad (1)$$

In most vision-based landing systems, it is assumed that the intrinsic matrix of the camera is precalibrated, whose focus does not change in flight. But in practical applications, its focus is to be changed by distant control for the purpose of shooting or inspection. With multiple-view vision technology, we could utilize multiple images in which target is included to estimate the intrinsic matrix of the camera (Hartley and Zisserman 2000; Zhang 2000) as shown in the following deduction:

Assume target is on  $Z = 0$  plane of the world coordinate system. The image coordinate of a pixel is denoted by  $\mathbf{m} = [u, v, 1]^T$ , and the world coordinate of three-dimensional point is denoted by  $\mathbf{M} = [x, y, z, 1]^T$ . According to projection theory, we have Equation (2):

$$\lambda \mathbf{m} = PM = \mathbf{K}[r_1, r_2, r_3, T] \begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} = \mathbf{K}[r_1, r_2, T] \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (2)$$

where  $\lambda$  is the unknown homogeneous scale factor,  $r_1, r_2, r_3$  is a row vector of rotation matrix, and  $T$  is the translator vector. Given  $\mathbf{H} = \mathbf{K}[r_1, r_2, T] = [h_1 \ h_2 \ h_3]$ , we can get Equation (3):

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{H} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3)$$

It is clear that  $\mathbf{H}$  is a homography between image plane and target plane. Provided

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{12} & h_{22} & h_{23} \\ h_{13} & h_{23} & h_{33} \end{bmatrix},$$

from Equation (3), we have,

$$h_{11}x + h_{12}y + h_{13} - h_{31}xu - h_{32}yu - h_{33}u = 0$$

$$h_{21}x + h_{22}y + h_{23} - h_{31}xv - h_{32}yv - h_{33}v = 0$$

Once a relation between the present image and the designed pattern is determined, given four or more such correspondences, we obtain a matrix equation of the form  $Ah = 0$ , where  $\mathbf{h}$  is just the components  $h_{ij}$  stacked into a nine-element vector. The solution  $h$  is the null space of  $A$ , which may be computed using the singular value decomposition (Hartley and Zisserman 2000; Yuan *et al.* 2004).

Using the knowledge that  $r_1$  and  $r_2$  are orthonormal, we have

$$\mathbf{h}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_2 = 0 \quad (4)$$

$$\mathbf{h}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_2 \quad (5)$$

Given

$$\mathbf{B} = \mathbf{K}^{-T} \mathbf{K}^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \quad (6)$$

Noted that  $\mathbf{B}$  is symmetric, defined by a six-dimensional vector  $\mathbf{b} = [B_{11}, B_{12}, B_{13}, B_{22}, B_{23}, B_{33}]^T$ . Therefore, the two fundamental constraints (4) and (5) can be rewritten as two homogeneous equations in  $\mathbf{b}$ :

$$\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \mathbf{b} = 0 \quad (7)$$

where  $V_1$  and  $V_2$  are six-dimensional vectors, only relative to elements of matrix  $\mathbf{H}$ . If  $n$  images of the model plane are observed, by stacking  $n$  such equations as Equation (7), we have

$$V \mathbf{b} = 0 \quad (8)$$

where  $V$  is a  $2n \times 6$  matrix. If  $n \geq 3$ , we will have, in general, a unique solution  $\mathbf{b}$  defined up to a scale factor. Once  $\mathbf{b}$  is estimated, we can compute all camera intrinsic matrix  $\mathbf{K}$ . Note that the movement of the camera for selected images must not be translating, or degenerated configuration will occur (Zhang 2000).

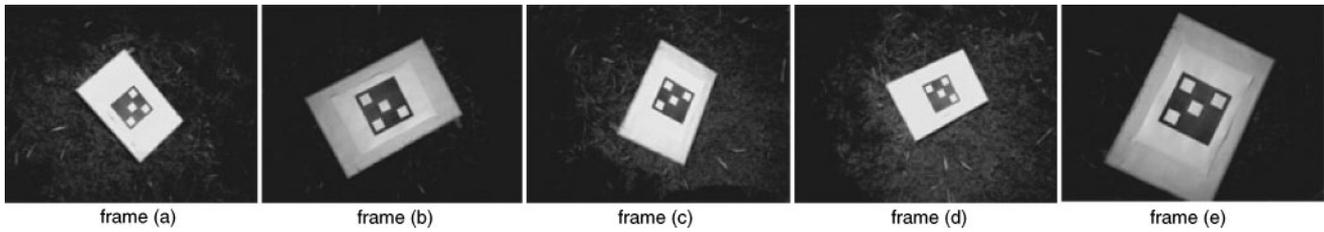


Figure 12 Five original images.

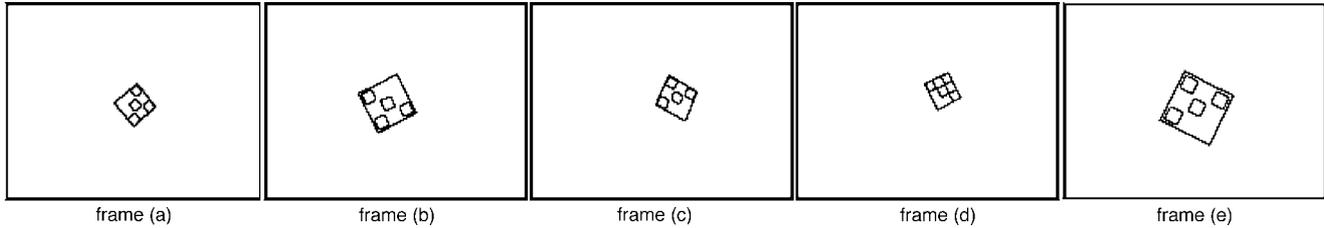


Figure 13 Five images after edge detection.

In practical application, we employ five pictures for calibration of the camera, and the five pictures are chosen through the following procedures: first, a target must be picked from the picture; besides, with large differences in cradle angle on shooting the five pictures, the angle can be obtained through a photoelectric encoder. During and after calibration, the focus of the camera can be fixed.

### POSE ESTIMATION

If the target is found in image, then we can get image coordinate  $\mathbf{m} = [u, v, 1]^T$  of 20 corners. With the already known world coordinator  $\mathbf{M} = [x, y, z, 1]^T$  for the 20 corners, according to the above calculation method,  $\mathbf{H}$  can also be obtained.

From (2) and (3), we have

$$[r_1, r_2, T] = \mathbf{K}^{-1}\mathbf{H} \quad (9)$$

where  $r_1, r_2$  and  $r_3$  are the column vectors of the rotation matrix. When  $\mathbf{K}$  and  $\mathbf{H}$  are known,  $r_1, r_2$  and  $T$  can be worked out from Equation (9). Since  $\mathbf{R}$  is a rotation matrix, namely,  $r_1, r_2$  and  $r_3$  are perpendicular,

$$r_3 = r_1 \times r_2 \quad (10)$$

Because of noise in the data, the so-computed matrix  $\mathbf{R} = [r_1, r_2, r_3]$  does not in general satisfy the properties of a rotation matrix. The problem considered is to solve the best rotation matrix  $\mathbf{R}$  to approximate a given  $3 \times 3$  matrix  $\mathbf{R}$ . Here, 'best' is in the sense of the smallest Frobenius norm of the difference  $\mathbf{R} - \mathbf{R}$ . That is,

$$\min \|\mathbf{R}' - \mathbf{R}\|_F \text{ subject to } \mathbf{R}'^T \mathbf{R}' = \mathbf{I}$$

$$\begin{aligned} \text{Since } \|\mathbf{R}' - \mathbf{R}\| &= \text{trace}(\mathbf{R}' - \mathbf{R})^T (\mathbf{R}' - \mathbf{R}) \\ &= 3 + \text{trace}(\mathbf{R}'^T \mathbf{R}) - 2 \text{trace}(\mathbf{R}'^T \mathbf{R}) \end{aligned} \quad (11)$$

Problem (11) is equivalent to one of maximizing  $\text{trace}(\mathbf{R}'^T \mathbf{R})$ . Let the singular value decomposition of  $\mathbf{R}$

be  $\mathbf{USV}^T$ , where  $\mathbf{S} = \text{diag}(s_1, s_2, s_3)$ . If we define an orthogonal matrix  $\mathbf{Z}$  by  $\mathbf{Z} = \mathbf{V}^T \mathbf{R}'^T \mathbf{U}$ , then

$$\begin{aligned} \text{trace}(\mathbf{R}'^T \mathbf{R}) &= \text{trace}(\mathbf{R}'^T \mathbf{USV}^T) = \text{trace}(\mathbf{R}'^T \mathbf{USV}^T) \\ \text{trace}(\mathbf{ZS}) &= \sum_{i=1}^3 z_{ii} s_i \leq \sum_{i=1}^3 s_i \end{aligned}$$

It is clear that the maximum is achieved by setting  $\mathbf{R}' = \mathbf{UV}^T$  because then  $\mathbf{Z} = \mathbf{I}$ . This gives the solution to (11).

*Note:* It is the pose of camera that is calculated here. As for the pose of UAV, cradle pose should also be taken into consideration at the same time, which will not be explained in detail here.

### EXPERIMENT

For protection of the UAV, a stable independent operation of the vision system should be ensured prior to its installation on the UAV. A ground simulation experiment was conducted to verify the effectiveness of the method. In our test, a laser printer was used to print the target, where the size of the black square is 180 mm  $\times$  180 mm; in practical application, the target size can be decided by the parameters of the camera. In testing, the camera moves (by hand) at a height of 500–2000 mm relative to the target. Polhemus fastrack six-dimensional magnetic tracker (with a static accuracy 0.8 mm RMS for the  $x, y$ , or  $z$  position; 0.15° RMS for receiver orientation) is installed on the camera, so as to distinctively reflect the computing result. To move the camera above the target and send image sequence wirelessly to computer for real-time processing, an image resolution of 320  $\times$  240 is used. If a high image resolution is chosen, more precise results will be obtained, but high transmission speeds cannot be reached. Frame number for the experiment is 120, and five images are automatically chosen from the first 40 frames for calibration with the method mentioned above, while taking into consideration

Table 1 Result of pose estimation

Frame	Translation (mm)	Rotation		
a	(-55.0, -223.3, 1062.8)	0.6633	0.7193	0.2064
		0.6899	-0.6947	0.2038
		0.2900	0.0072	-0.9570
b	(257.76, 63.09, 782.07)	-0.8586	0.4832	-0.1713
		0.4715	0.8755	0.1059
		0.2011	0.0101	-0.9795
c	(-298.68, 365.40, 862.86)	0.3782	-0.8814	0.2831
		-0.8464	-0.4531	-0.2798
		0.3748	-0.1338	-0.9174
d	(-43.69, 55.60, 1042.6)	0.7949	-0.4787	0.3727
		-0.4796	-0.8721	-0.0972
		0.3716	-0.1015	-0.9228
e	(-2.54, 106.79, 630.72)	0.3910	-0.9126	0.1191
		-0.9059	-0.4045	-0.1256
		0.1628	-0.0588	-0.9849

the angle signal of the tracker, and the calibration result is

$$f_x = 293.75, \quad f_y = 288.76, \quad \gamma = 0.0, \\ u_0 = 150.66, \quad v_0 = 142.98$$

Following this, the pose estimation can be done. We chose five frames randomly (see Figure 12), and the result of pose estimation is given in Table 1 and Figure 14.

Error occurred in five frames when detecting the target in the other 80 frames, so we have an accuracy of 93.75%; its processing speed reaches 25 frames per second. Average absolute value of divergence compared to measurement of the magnetic tracker is as follows:

$$|\Delta x| = 6 \text{ mm}, \quad |\Delta y| = 6 \text{ mm}, \quad |\Delta z| = 8 \text{ mm}, \\ |\Delta \text{roll}| = 2.5^\circ, \quad |\Delta \text{pitch}| = 3.1^\circ, \quad |\Delta \text{yaw}| = 2.7^\circ$$

Detection errors occur when the angle between the ray axis of the camera and the target plane is large, or the images of the target have too serious distortion. But autonomous landing will not be affected distinctively. During autonomous landing, the UVA is in the state of slow speed and acceleration; so apparent errors can be got rid of by checking motion continuity. Besides, since there is a fast refresh rate of the system currently, small errors will not affect the UVA for a long time, and thus can be ignored. For protection from any danger, an import control rule is that, in case any signal is lost or there is an error, our UAV will be in the protective mode, in which case the UAV will automatically hold its position and pose in air.

## CONCLUSION

The real-time vision-based method for guided autonomous landing of rotor-craft unmanned aerial vehicle introduced here has two features:

1. It combines designing and identification of the target, which simplifies image processing, thus improving its real-time effect and accuracy.

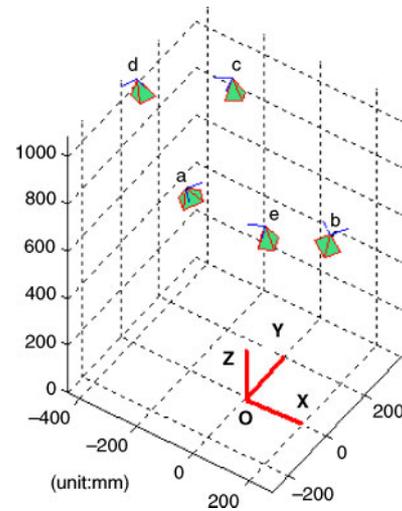


Figure 14 The pose of five frames present in world coordinates.

2. It adopts multiple-view vision technology and combines angle sensor to calibrate the intrinsic parameters of the camera online, thus improving its flexibility and practicality.

Although this is done in linear computing, it is accurate enough to guide an autonomous landing of the UAV; it has also been proved to be a robust method through the theoretical analysis and experiment.

Next, debugging is to be done on the mini rotor-craft UAV; based on motion prediction, cradle heads are to be controlled to track the landing target; still, we are to search for suitable hardware on UAV to perform a vision job with light weight and low power.

## ACKNOWLEDGMENT

This research was supported by the National High Technology Research and Development Program of China (863 Program) under the grant number 2004AA420110.

## REFERENCES

- Gong Z. 2003. SUAVs for firefighting and rescuing. In IARP on Micro Robots, Micro Machines and Micro Systems, April 2003.
- Hartley R, Zisserman A. 2000. *Multiple View Geometry in Computer Vision* [M]. London: Cambridge University Press.
- Rekimoto J. 1998. Matrix: A realtime object identification and registration method for augmented reality. In 3rd Asia Pacific Proceedings of Computer Human Interaction, pp. 63–8.
- Saripalli S, Montgomery J. 2002. Vision-based autonomous landing of an unmanned aerial vehicle. In IEEE International Conference on Robotics and Automation, Washington DC, pp 2799–804.
- Saripalli S, Montgomery JF, Sukhatme GS. 2003. Visually-guided landing of an unmanned aerial vehicle. *IEEE Trans Robotics Autom*, 19(3):371–81.

- Saripalli S, Sukhatme GS. 2003. Landing on a moving target using an autonomous helicopter. In Proceedings of the International Conference on Field and Service Robotics, July 2003.
- Shakernia O, Ma Y, John Koo T, Shankar Sastry S. 1999. Landing an unmanned air vehicle vision based motion estimation and non-linear control. *Asian J Control*, 1:128–45.
- Sharp C S, Shakernia O. 2001. A vision system for landing an unmanned aerial vehicle. In Proceedings of IEEE ICRA, pp. 1720–7.
- Shakernia O, Vidal R. 2002. Multiple view motion estimation and control for landing an unmanned aerial vehicle. In IEEE ICRA, Washington, DC, pp 2793–8.
- Yuan Z, Gong Z, Wu J. 2004. A registration method of AR system applicable to outdoor environment [C]. In ICIMA 2004 Conference, August 2004, Chengdu, People's Republic of China, pp. 901–4.
- Zhang Z 2000. A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell*, 22(11):1330–4.





**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

