

Attitude Estimation Algorithm and Verification based on Binocular Vision of Micro Aerial Vehicles

Yongqin Wan^{1, a}, Xi Zhang^{1, b} and Lijing Li^{1, c}

¹School of Instrumentation Science & Opto-electronics Engineering, Beihang University, Beijing 100191, China.

^asusan2012_ok@126.com, ^bzhangxi@buaa.edu.cn, ^clilj@eyou.com

Keywords: Micro Aerial Vehicles (MAVs), Binocular Vision, Attitude Estimation.

Abstract. Aiming at the navigation of Micro Air Vehicles (MAVs) in the natural environment, an algorithm is studied to measure the attitude of MAVs based on binocular vision. Binocular images are taken by calibrated binocular camera. The image feature points are extracted by Harris algorithm, feature points are matched with Normalized Cross Correlation (NCC) algorithm, and obtained a set of matching points. The three dimensional coordinates of the feature points can be calculated in camera coordinate system. Then the attitude of MAVs is estimated by Singular Value Decomposition (SVD) method and Least Median Squares (LMedS) theorem. The experiment result shows that the attitude estimation error is small, which can meet the requirements of MAVs navigation.

1. Introduction

Micro Air Vehicles (MAVs) have the characteristics of small size, light weight and high security. They are simple and easy to conceal when being used, having a wide application prospects both in military and civil fields^[1]. With the development of computer vision theory, visual navigation has become a research hotspot in the field of MAVs navigation^[2].

Binocular vision uses two relatively fixed cameras to obtain the images of the objects at once, then calculates position deviation of the images in different camera to make 3D reconstruction, the attitude information can be calculated according to the reconstructed 3D coordinates^[3]. In the early period, it is mainly applied to the navigation based on feature mark and a priori natural scene^[4,5]. However, it has limitations in the practical application. In this paper, binocular vision is used to measure MAVs' attitude in the natural environment.

2. The Principle of MAVs' Attitude Estimation with Binocular Vision

The principle of binocular vision measuring MAVs' attitude is: feature points extraction and stereo matching are carried out respectively in stereo images of front and back moment, then common feature points of front and back moment are matched in the two matching point sets. Finally, the attitude can be calculated according to the reconstructed 3D coordinates of common feature points.

The camera imaging model is approximate to the pinhole model. Define pixel coordinate system (u, v) , image coordinate system (x, y) , camera coordinate system (X_c, Y_c, Z_c) and world coordinate system (X_w, Y_w, Z_w) , as shown in Fig. 1. The X_c axis parallels to u axis, Y_c axis parallels to v axis, and Z_c axis coincides to the optical center. According to the relationship of perspective projection, the relationship of any point M and pixel coordinate system can be got:

$$\begin{aligned}
Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = KM \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = P \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
\end{aligned} \tag{1}$$

Among them, $(u, v, 1)^T$ and $(X_w, Y_w, Z_w, 1)^T$ are homogeneous coordinate value of the characteristic points in pixel coordinate system and world coordinate system. \mathbf{K} is the camera intrinsic matrix, \mathbf{M} is the extrinsic camera matrix, \mathbf{R} and \mathbf{t} are the rotation matrix and translation vector of world coordinate system relative to camera coordinate system; \mathbf{P} is a projection matrix.

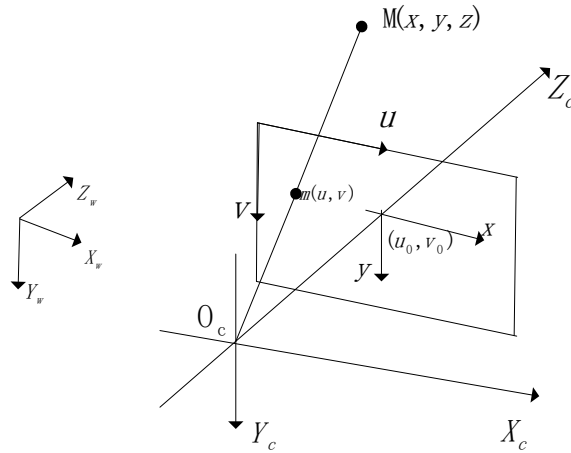


Fig. 1 Relationships between coordinate systems

2.1 Feature Points Extraction and Stereo Matching.

In order to ensure the accuracy of stereo matching, the extracted feature points should be sufficiently clear and widely distributed. Corner points are extracted in this paper. The algorithm of corner points extraction have Moravec^[6], Harris^[7], SIFT^[8], et al. The extracted corner points are unstable with Moravec algorithm if images have large noise. The computation of SIFT algorithm is more complicated, and the real-time performance is poor. So the feature points are detected with Harris algorithm.

After extracted feature points, the feature points are matched. To realize 3D reconstruction, the feature points in the left and right images are matched. Similarly, To realize attitude estimation, the feature points in the front and back frame images are matched.

Attitude estimation requires at least 4 images of two cameras in front and back moments. A period of stereo matching is divided into the following three steps. Firstly, the feature points of previous moment images are matched, left image as a benchmark, and the corresponding feature points in right image are searched. Because of left and right images are rectified, the feature points matching limits on horizontal axis. Secondly, the feature points of current moment images are matched, and method is same as first procedure, which obtained the matching point set. Finally, common feature points of previous and current moments are searched in the two matching point sets. The matching point set of current moment as a benchmark, the corresponding points are searched in the previous moment point set. Feature points are matched with Normalized Cross Correlation (NCC) algorithm^[9]. The gray value of image is normalized, the correlation coefficient is calculated according to gray information around the feature points. The larger the correlation coefficient, the higher the matching degree is. In this paper, the correlation coefficient is calculated in 5*5 window size.

2.2 3D Reconstruction.

Fig. 2 as a general type of binocular vision model. The world coordinate system $O-XYZ$ is built on left camera and right camera coordinate is $O_c^r - X_c^r Y_c^r Z_c^r$, point $M(X, Y, Z)^T$ in left and right pixel coordinate system is m_l and m_r .

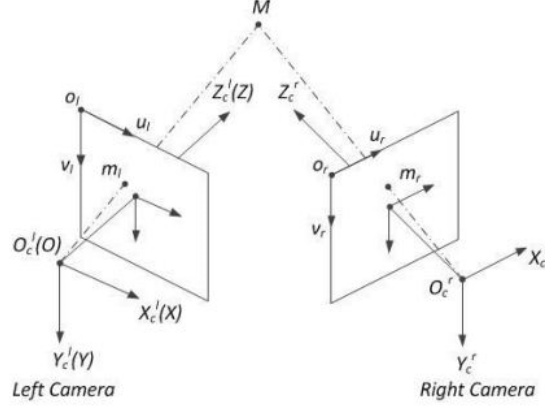


Fig. 2 General type of binocular vision model

Assuming that K_l, K_r are the intrinsic matrix of left and right camera respectively, meanwhile P_l, P_r are the project matrix of left and right camera respectively. The transform relationship between two cameras' coordinate system represents by rotation matrix R and translation vector t as follows:

$$z_l m_l = K_l [I \ 0] M = P_l M \quad (2)$$

$$z_r m_r = K_r [R \ t] M = P_r M \quad (3)$$

The linear equations are obtained as:

$$\begin{bmatrix} u_l p_{31}^l - p_{11}^l & u_l p_{32}^l - p_{12}^l & u_l p_{33}^l - p_{13}^l \\ v_l p_{31}^l - p_{21}^l & v_l p_{32}^l - p_{22}^l & v_l p_{33}^l - p_{23}^l \\ u_r p_{31}^r - p_{11}^r & u_r p_{32}^r - p_{12}^r & u_r p_{33}^r - p_{13}^r \\ v_r p_{31}^r - p_{21}^r & v_r p_{32}^r - p_{22}^r & v_r p_{33}^r - p_{23}^r \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{14}^l - u_l p_{34}^l \\ p_{24}^l - v_l p_{34}^l \\ p_{14}^r - u_r p_{34}^r \\ p_{24}^r - v_r p_{34}^r \end{bmatrix} \quad (4)$$

The values of X, Y , and Z can be obtained by solving equation (4), which use least square method^[10].

2.3 Attitude Estimation.

2.3.1 Initial Attitude Estimation.

Supposing that the coordinate value of n known space characteristic points in world coordinate system is $X_{wi} = (X_{wi}, Y_{wi}, Z_{wi})^T, i=1 \dots n$. After capturing images with binocular camera, the coordinate of feature points in camera coordinate system can be obtained by feature points extraction, stereo matching and 3D reconstruction. Continually, initial motion parameters R and t can be calculated by the coordinate value of camera coordinate system and world coordinate system.

2.3.2 Motion Attitude Estimation.

To the left camera as a benchmark, feature points in the previous frame and current frame images have n stereo matching pairs, indicating that M_{pi} and $M_{ci} (i=1, 2, \dots, n)$. Thus their relationships can be expressed as:

$$M_{ci} = R_p M_{pi} + t_p \quad (5)$$

According to the formula (5), the performance function is defined as follows:

$$E(R, t) = \sum_{i=1}^n \|M_{ci} - (R_p M_{pi} + t_p)\|^2 \quad (6)$$

In addition, the rotation matrix R must be the unit orthogonal matrix, $r_i^T r_i = 1, r_i^T r_j = 0, i, j=1, 2, 3, i \neq j$. In this paper, the Singular Value Decomposition (SVD) method^[11] is chosen to solve nonlinear

least squares problems. Meanwhile the Least Median Squares (LMedS) theorem is used to overcome the influence of outliers. Set k_i for the residual error of the i^{th} , thus

$$k_{pi} = \|M_{ci} - (R_p M_{pi} + t_p)\| \quad (7)$$

The Least Median Squares (LMedS) theorem estimates motion parameters by solving the following nonlinear optimization problems.

$$(\hat{R}, \hat{t}) = \min \text{med}_i k_i^2 \quad (8)$$

Among them, k_i^2 is regression variance, $\text{med}()$ is median operation.

Therefore, m samples are randomly selected in n matching feature points, each sample is composed by q (in this paper $q=3$) non-collinear points; Taking each sample point into SVD method to calculate $E_p(R, t)$; then calculate regression variance of the matching feature points to find median $M_p = \text{med}_{i=1 \dots n} k_i^2$, and the minimum M_p corresponding to the $E_p(R, t)$ is the best parameter of motion estimation. Thus the relative attitude can be calculated by the following formula.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (9)$$

$$= \begin{bmatrix} \sin \phi \sin \theta \sin \gamma + \cos \phi \cos \gamma & \cos \theta \sin \gamma & \cos \phi \sin \theta \sin \gamma - \sin \phi \cos \gamma \\ \sin \phi \sin \theta \cos \gamma - \cos \phi \cos \gamma & \cos \theta \cos \gamma & \cos \phi \sin \theta \cos \gamma + \sin \phi \sin \gamma \\ \sin \phi \cos \theta & -\sin \theta & \cos \phi \cos \theta \end{bmatrix}$$

Among them, θ is pitch angle, r is roll angle, ϕ is yaw angle. And Attitude angle is

$$\begin{cases} \theta = \arcsin(-r_{32}) \\ r = \arctan(r_{12} / r_{22}) \\ \phi = \arctan(r_{31} / r_{33}) \end{cases} \quad (10)$$

The Attitude of MAVs is relative to world coordinate system can be obtained by the superposition after obtaining initial Attitude and relative rotation matrix.

3. Experiment Result and Analysis

In order to verify the correctness of algorithms, experiment is carried out with binocular camera which consists of 2 MER-200-14Gx Industrial cameras. The cameras use CCD chips, the frame rate of camera is 14fps and the pixel size is 4.4um*4.4um. The cameras are calibrated before experiment. Calibration parameters are shown in Table 1.

Table 1 Binocular Camera Calibration Parameters

Type	Parameter	Value
Left camera	Focal distance (fc_left)	[3982.09 3981.33] \pm [9.24 8.74]
	Principal point (cc_left)	[815.93 626.48] \pm [14.21 14.84]
Right camera	Focal distance (fc_right)	[3989.24 3989.34] \pm [6.60 6.58]
	Principal point (cc_right)	[835.20 602.04] \pm [11.35 11.78]
Extrinsic parameter	Rotation matrix	om = [-0.0042 0.5303 0.0169] \pm [0.0031 0.0042 0.0010]
	Translation vector	T = [-436.38 -3.35 114.44] \pm [0.60 0.33 2.01]

Because the algorithms are in verification stage, off-line verification is carried out in this paper. The image processing programs are running on PC, and the attitude estimation flow chart is shown in Fig.3.

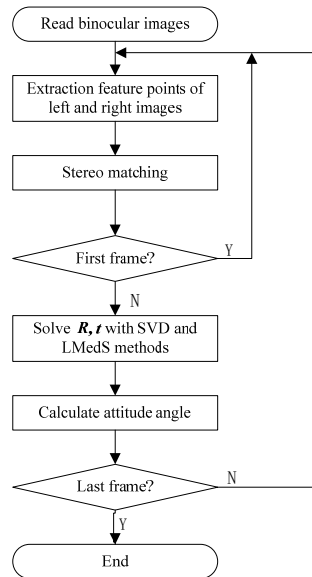


Fig. 3 Flow chart of attitude estimation of MAVs

The binocular camera is moved around x axis by 15 deg counterclockwise, and 2 stereo pairs are taken by moving the camera back and forth. After Feature points extraction and stereo matching, 4 common feature points are obtained, and calculate the coordinates of camera coordinate system. 3 non-collinear feature points are randomly selected to calculate rotation matrix R , which is used to estimate attitude as shown in Table 2.

Table 2 Attitude Estimation Value

Numble	Pitch Angle (deg)	Roll Angle (deg)	Yaw Angle (deg)
1	15.79	0.43	0.16
2	15.77	0.43	0.19
3	15.80	0.45	0.17
4	15.83	0.46	0.18

The RMSE of attitude is [0.80, 0.44, 0.18] (deg). Compared to the reference [12], the error is relatively small, which can meet the requirements of visual navigation. Also, general binocular model is used to calculate 3D reconstruction in this paper, has a practical significance.

4. Conclusions

The attitude estimation of MAVs based on binocular vision is studied, and a kind of algorithm is described. The experiment result shows that attitude estimation error is small, which provides a new technique for the navigation of MAVs. Visual navigation and inertial navigation can be integrated in the future.

References

- [1]. Haijun Shen, Xu Wang. Artificial insect-MAV [J]. Science & Technology Vision. Vol.1 (2014) No. 1, p. 26-29.
- [2]. Yongjun Yu, Jinfa Xu, Liang Zhang, et al. Research on SINS/binocular vision integrated position and attitude estimation algorithm [J]. Chinese Journal of Scientific Instrument. Vol. 35 (2014) No. 10, p. 2170-2176.
- [3]. Songde Ma, Zhengyou Zhang. Computer vision- computing theory and algorithm foundation [M]. Science Press, 2003, p. 173-178.

- [4]. Saripalli S., Montgomery J. F., Sukhatme G. S.. Vision-based autonomous landing of an unmanned aerial vehicle [A]. Proceedings 2002 IEEE International Conference on Robotics and Automation [C], Washington, D. C. , USA, 2002, p. 2799-2804.
- [5]. Xiang Pan, Wanwan Tong, Zhesheng Jiang. Runway detection and tracking for UAV visual navigation [J]. Chinese Journal of Sensors and Actuators. Vol. 23 (2010) No.6, p.820-824.
- [6]. Moravec H. P.. Towards automatic visual obstacle avoidance [A]. Proceedings of the 5th International Joint Conference on Artificial Intelligence [C], Manchester, England, 1977, p. 584-586.
- [7]. Harris C., Stephens M.. A combined corner and edge detector [A]. Proceedings of Fourth Alvey Vision Conference [C]. Manchester, England, 1988, p. 189-192.
- [8]. Qi F., Weihong X., Qiang L. Research of Image Matching Based on Improved SURF Algorithm [J]. Telkomnika Indonesian Journal of Electrical Engineering. Vol. 12 (2014) No. 2, p. 1395-1402.
- [9]. Xiaochuan Zhao. Matlab image processing- the ability to improve and application case [M]. Science press, 2014, p. 51.
- [10]. Guangjun Zhang. Machine Vision [M]. Beijing: Science press, 2008, p. 105-108.
- [11]. Wei Hong: Position and attitude estimation techniques of MAV based on vision/MEMS [D]. Master Degree, Harbin Institute of Technology, China, 2010. p.44.
- [12]. Liang Zhang, Jinfu Xu, Qingyuan Xia. Pose estimation algorithm and verification based on binocular stereo vision for unmanned aerial vehicle [J]. Journal of Harbin institute of technology, Vol. 46 (2014) No.5, p.66-72.