Optical flow based 3D motion estimation for autonomous landing an UAV on deck

Wang Rui Zhang Guangjun Yan Peng
(School of Instrumentation and Optoelectronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

ABSTRACT

3-D motion estimation method based on computer vision theory is employed to implement a vision guide algorithm for UAV in this paper. First, the image sequences of landing target are taken by the camera mounted on UAV with known focal length and the Lucas-Kanade method is adopted to estimate two successive frame optical flow; then a hierarchical approach is described to effectively decompose the nonlinearities of the 3-D motion estimation into two linear subsystems; finally 3-D motion and structure(depth) information of landing target relative to UAV is recovered without using features of landing target . Experiments using both computer simulated images and real video images demonstrate the correctness and effectiveness of our method.

Keywords: Optical flow, 3D motion estimation, vision guiding, Unmanned aerial vehicle

1 VISION-GUIDING PRINCIPLE FOR UAV AUTO-LANDING ON DECK

Unmanned Aerial Vehicles (UAV), particularly ones with vertical takeoff and landing capabilities (VTOL), enable to prevent harmful events for human pilots. A VTOL UAV with capability of autonomous landing on a ship would be very useful to perform various tasks, such as search, rescue, and other military scenarios. An autonomous landing maneuver on a ship in the terminal period depends largely on two capabilities: the decision of when to land and the generation of control signals to guide the UAV to a safe landing. This paper focuses on the two capabilities by adopting traditional vision-guiding approaches relying on a high-contrast planar landing target on a ship, that can be easily identified by a nadir-pointing monocular camera mounted on the UAV. The problem addressed here is a special case of 3D motion estimation between on-board camera and the ship since all feature points lie on a planar surface (the landing target). 3D motion estimation can be roughly classified as feature-based or flow-based, according to whether the data they use are a set of feature matches or an optical flow field. We can recover 3-D motion parameters using a set of 2-D images, only computing the optical flow between two successive images, and get the relative position and orientation of the UAV.

We make the following assumptions which simplify the problem but are not restrictive. A monocular camera is assumed being fixed to the UAV and the optical axis of the camera coincides with the vertical axis of the UAV body frame. In the terminal period of autonomous landing, the UAV is near hover at all instances over the high-contrast planar landing target on the deck and descends gradually. Then the UAV waits its chance to land in the area of landing target on the basis of the criterion: Relative to the UAV, the deck must be in the state of descend with maximum velocity less than 2m/s in the vertical direction and its rotation angles less than 5°. From sequence of images for landing target taken by the on-board camera with known focal length *f*, the 3D motion Parameters between the camera and the landing target can be

International Conference on Space Information Technology, edited by Cheng Wang, Shan Zhong,
Xiulin Hu, Proc. of SPIE Vol. 5985, 59854H, (2005) · 0277-786X/05/$15 · doi: 10.1117/12.658348

related through the optical flow, and the schematic diagram of vision-guiding for UAV based on optical flow is illustrated in Fig.1. Particularly, our work in this paper is focused on the part of rectangles with full line in Fig.1.

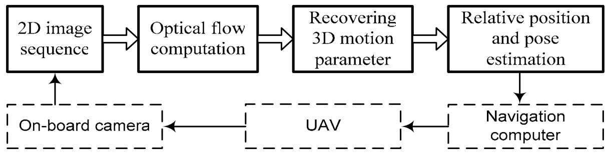


Fig. 1 Schematic diagram of vision-guiding for UAV

2 OPTICAL FLOW BASED MOTION ANALYSIS

A key task for any vision system is to extract information from a sequence of time varying images. This exercise is called Motion Analysis. *Optical flow* is one of the important tools in Motion Analysis and is formally defined as the *apparent motion of image brightness patterns in an image sequence.* It is an approximation of the perspective projection of a set of 3-D motion vectors, known as the *Motion Field*, onto an image plane. The motion field, and hence its approximation, the optical flow, carries rich information regarding the 3-D motion parameters.

2.1 Optical flow computation

Let $I(x, y, t)$ be the intensity value of the image at pixel (x, y) in time t. If the immediate neighborhood of (x,y) is translated some very small distance (dx, dy) during the interval dt; the optical flow gradient constraint equation should be satisfied:

$$I_x u + I_y v + I_t = 0 \tag{1}$$

Where $u=\dot{x}$ and $v=\dot{y}$, are components of optical flow in horizontal and vertical directions respectively; I_x, I_y and I_t denote partial derivatives of $I(x, y, t)$ with respect to x, y and t respectively. Comparison of performance of many optical flow techniques, local differential approaches (Lucas and Kanade) were found to be most accurate and robust[1].

Following Lucas and Kanade[2], we implement a weighted least-squares fit of equation (1) to a constant model for $(u,v)^T$ in each local region Ω by minimizing

$$\sum_{X\in\Omega} W^2(X)(I_x u + I_y v + I_t)^2 \tag{2}$$

Where $W(X)$ is a window that gives more influence to constraints near the center of the neighborhood Ω. The solution to (3) is given by

$$\begin{bmatrix} \sum W^2(X) I_x^2(X) & \sum W^2(X) I_x(X) I_y(X) \\ \sum W^2(X) I_x(X) I_y(X) & \sum W^2(X) I_y^2(X) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -\sum W^2(X) I_x(X) I_t(X) \\ -\sum W^2(X) I_y(X) I_t(X) \end{bmatrix} \tag{3}$$

This can be reduced to a closed form solution for the flow estimates, viz., optical flow vector $(u,v)^T$.

2.2 Camera motion and imaging models[2,3,5]

To recover the motion of the camera using image measurements of some points in the landing target from the so called optical flow estimation, we should formulate both the camera motion and the depth Z of the landing target derived from the optical flow $(u,v)^T$. The imaging of the camera is given by the pin-hole model perspective projection of points in the 3D world onto the image plane, the camera coordinate is set as a three-dimensional Cartesian coordinate system with the origin at the camera's lens and z-axis lying along the direction of the optical axis of the camera, pointing downward, and that y-axis and x-axis run parallel to the longitudinal axis and lateral axis of the UAV respectively. The camera coordinate (X,Y,Z) of a point in the space and the image plane coordinate (x, y) of its image are related by the perspective transformation as[7]: $$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \tag{4}$$

Without loss of generality, we assume that the rotation angles of the oscillating deck are small. Appealing to the theory of the rigid body motion, the velocity of the arbitrary point *(X,Y, Z)* and the depth *Z* in the landing target can be related through the optical flow$(u,v)^T$:

$$\begin{cases} u = \frac{1}{Z}(ft_x - xt_z) - \frac{1}{f}xy\,\omega_x - y\omega_z + (f + \frac{x^2}{f})\omega_y \\ v = \frac{1}{Z}(ft_y - yt_z) + \frac{1}{f}\omega_y xy - f\omega_z - (f + \frac{y^2}{f})\omega_x \end{cases} \tag{5}$$

In which, $T=(t_x, t_y, t_z)^T$ and $R=(\omega_x, \omega_y, \omega_z)^T$ is the transition and the rotation between two sequence images, or

in matrix form: $$V = \hat{Z}AT + BR \tag{6}$$

Where $V = [u \quad v]^T$ is the flow vector, (*x, y*) are image coordinates, and

$\hat{Z} = \frac{1}{Z}$, $A = \begin{bmatrix} f & 0 & -x \\ 0 & f & -y \end{bmatrix}$, $B = \begin{bmatrix} -xy/f & f + x^2/f & -y \\ -(f + y^2/f) & xy/f & x \end{bmatrix}$

2.3 3D motion parameters from hierarchical framework

A challenging problem to recover (*T,R)* by solving the nonlinear system (6). In addition, algorithms based on derivatives of optical flow are very sensitive to noise in the optical flow estimates. Nevertheless, the observation can be obtained from the view of equation (6): If *Z* is known, (6) is a linear system with the motion parameters *(T, R)* unknown; If *(T, R)* is known, (6) is a linear system with the depth *Z* unknown. These two linear subsystems enable us to develop a novel solution technique of a hierarchical framework .Then due to the fact that the dynamic depth variation for the surface of the coplanar target is small compared to the distance between the optical center and the target, a hierarchical iterative algorithm is introduced as follows:

After dividing both sides in (6) by 2^n, where the integer *n* is the level index, equation (6) can be written as

$$V_n = \hat{Z}_n A_n T + B_n R \tag{7}$$

The definitions of (x_n, y_n, Z_n) and (u_n, v_n) define a hierarchical "scene", in which the sizes of all objects and the optical flow are reduced by half when the level *n* increases by one. The nonlinear system of motion, depth, and flow is defined by (7) for each level of such a hierarchical "scene". Based on above deductions, the nonlinear system (6) is solved through the following two linear subsystems from coarse to fine: $$S_n M = V_n, S_n = [\hat{Z}A_n \quad B_n], \quad M = [T \quad R]^T \tag{8}$$

$$E_n\hat{Z} = F_n, \quad E_n = A_n T, \quad F_n = V_n - B_l R \tag{9}$$

The motion (*T* , *R*) is estimated from (8) with Z_n at the maximum level initialized to a constant. The resulting is then passed to (9) and $\hat{Z}_n$ is re-computed. These estimations are repeated at each level until the motion is stableormaximum number of iterations is reached. $\hat{Z}_n$ is then propagated to the next finer level and refined, until $\hat{Z}_0$ or Z is solved.

2.4 Summary

The recommended hierarchical algorithm with two linear subsystems is as follows:

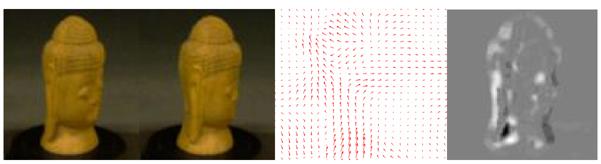
To establish the optical flow; The local differential method is used in the optical flow computation. Spatial neighborhood were 5×5 pixels, window function ***W(X)*** was separable and isotropic; its 1-d weights in the horizontal and vertical direction were (0.0625 0.25 0.375 0.25 0.0625) .Both vector and intensity fields of optical flow were obtained in this step;To smooth the optical flow intensity field with 3×3 median filter; Using hierarchical iterative algorithm, the depth and the velocity of the arbitrary point (X,Y,Z) in the landing target can be recovered.

3 EXPERIMENTAL RESULTS

The effectiveness and the performance of our hierarchical approach to 3D motion estimation from two-frame optical flow have been tested using both computer simulated dada and the real video data.

3.1 Computer simulated Experiments

The source data, including two frames of “Buddha” image sequence as shown in Fig. 2 (a) and corresponding texts of motion parameters, were downloaded from http://sampl.eng.ohio-state.edu/ ˜sampl/ data/. All frames are 200 × 200 pixels. The image sequence were generated from a virtual camera with the ground truth focal length (*f=1000* pixel) and motion parameters.



(a) (b) (c)

Fig. 2 (a) Two frames of “Buddha” image sequence
(b) Optical flow vector (left head part)
(c) Optical flow intensity distribution

Tab. 1 Estimation of 3D motion parameters

Motion parameters	Real value	Computation value	
		(35,28)	(51,33)
t_x/pixel	80	112.56	118.74
t_y/pixel	-110	-75.8	-78.41
t_z/pixel	-40	-81.43	-75.33
ω_x/rad	-0.110	-0.123	-0.121
ω_y/radω_z/rad	0.110	0.121	0.122
	0	0.000848	0.00132

Following the computation described in section 2.4, we can get the results including the optical flow vectors shown in Fig. 2(b) (left head part), the optical flow intensity distribution in Fig. 2(c), and also the parameters *(T,R)*for the image, e.g. image points: (35,28) and (51,33) (cf. table 1). In order to calculate the accuracy of depth *Z* estimation, the relative error for each 30 sample points is calculated and displayed inFig. 3.

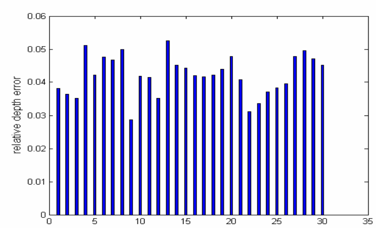


Fig.3 Distribution of depth relative error

Table 1 shows that the estimation error is about 40 pixels in transition and about 0.01° in rotation. Together with the fact that the relative error of depth estimation is less than 5.5% (cf. Fig.3), our hierarchical approach is demonstrated as a correct and effective method.

3.2 The Experiment with real video data

The experiment set-up is composed by a cradle head which can pan and tilt with the uniform velocity of 0.208 rad/s and 0.103 rad/s respectively, and a camera with the focal length of 12 mm viz. 2000 pixels. The image sequences of a high-contrast featured landing target is taken with the camera mounted on the moving cradle head, which is Pal TV signal format with the rate of 25 frames/s , the image resolution is 768×576. These image sequences can be simulated as that of the relative motion between UAV and the deck. Our implementation first smoothes any two successive frames image sequence with an isotropic spatial low-pass Gaussian filter with a standard deviation of 1.5 pixels, as shown in Fig.4(a). Then we can estimate two-frame optical flow from, the results are displayed in Fig.4(b). Moreover, the relative dynamic 3D structure *(R,T and Z)* of four selected points (every three points are not on a line) in the landing target can be extracted by the computation described in section 2.4. Appealing to the principle of space analytic geometry, the rotation angle θ between the deck and the image plane can be derived from 3D information at time *t1* and *t2* (cf. table 2 and table3.

(3).

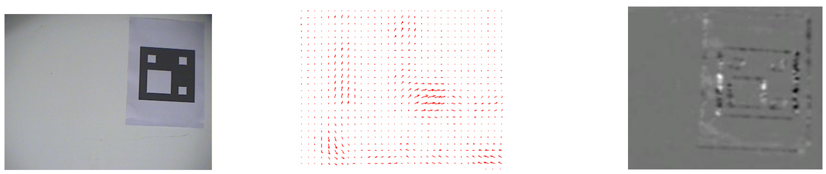


Fig.4 (a) One frame of target sequence; (b)Optical flow vector (local); (c)Optical flow intensity distribution

Tab. 2 Estimation of the rotation angle θ at $t1(\theta_{real} = 8^\circ)$

3D coordinates of three points			**Computation Value** θ/°	
X/pixel	Y/pixel	Z/pixel		
163	254	163125	7.3825	>5°
244	368	163637	Conclusion:	
255	205	164113	**Landing impossible**	

Tab. 3 Estimation of the rotation angle θ at $t2(\theta_{real} = 4^\circ)$

3D coordinates of three points			**Computation Value** θ/°	
X/pixel	Y/pixel	Z/pixel		
56	164	162452	3.8163	<5°
158	272	162857	Conclusion:	
231	58	163315	**Landing possible**	

Table 2 and table 3 show that not only the relative descend velocity t_z but also the rotation angle θ can be derived with the hierarchical representation. Experiment with real video data also illustrates that the absolute errors for the estimation of the relative rotation angle are less than 1° when the real oscillating angle less than 10°. So, the proposed approach is availability in the moderate sea state [4].

4. CONCLUSION

Flow-based 3-D motion estimation method is employed to implement a vision guide algorithm for UAV in the terminal period of autonomous landing on deck. A hierarchical representation for the relationship between the optical flow, depth, and the motion parameters is derived, and the resulting non-linear system is iteratively solved through two linear subsystems. The advantage of our techniques as opposed to feature-based techniques is that a repeat of segmentation of images and feature matching are not required. Meanwhile, the time aliasing problem is alleviated at the coarse level for any two-frame optical flow estimate so that the 3D motion estimation tends to be more accurate.

The experiment with both computer simulated data and the real video data indicate the effectiveness and availability of the proposed approach. From the error analysis, we know that in order to get more robust and accurate 3D motion estimation, the improvement for optical flow estimation and the research for strategy how to select the points on the target should be done in the future.

Acknowledgement: This work is supported by Aeronautical Science Fund, No.03I51009 &. National Science Fund for Distinguished Young Scholar, No.50125518

REFERENCES

1. J.L.Barron, D.J.Fleet and S.S Beauchemin "Performance of optical flow techniques"[J], Int.J.Computer Vision ,1994 ,12(1):333-340

2.B.Lucas and T.Kanade "An iterative image registration technique with an application to stereo vision"[A]. Proceedings of DARPA IU Workshop, 1981，page 121-130

3.Haiying Liu, Rama Chellappa "A hierarchical approach for obtaining structure from two-frame optical flow"[A]. Proceeding of Motion and Video Computing Workshop. 2002，page 214-219

4.Yang Peng, Wang Rui, Zhang Guang-jun, Zhou Fu-qiang "Software Imitation for Unmanned Aerial Vehicle Landing Scenes" [J]. Journal of Beijing University of Aeronautics and Astronautics, 2005，31(6)：609-613 (in Chinese).

5. Jia Yunde. *Machine vision*[M]. Beijing: Science Press, 1999, 60~72 ((in Chinese)