

## 小型无人机立体视觉目标追踪定位方法

王亭亭<sup>1</sup>, 蔡志浩<sup>1,2</sup>, 王英勋<sup>1,2</sup>

(1.北京航空航天大学,北京 100191; 2.飞行器控制一体化技术重点实验室,北京 100191)

**摘要:** 小型无人机在无卫星导航条件下使用的需求日益强烈,采用立体视觉方法可满足无人机在复杂环境中的目标检测、跟踪与定位。研究了双目立体视觉系统标定、特征点检测提取与匹配、双目测距等技术,实现了对图像中的指定目标进行追踪,同时实现对目标的定位,得到目标的位置,视觉系统标定误差在0.2个像素点以内,目标定位误差小于10 cm。以地面机器人作为追踪目标进行算法测试验证,在目标消失时间较短的情况下,能够达到较好的追踪效果。

**关键词:** 无人机; 双目视觉; 目标追踪; 摄像机标定; CamShift 算法

**中图分类号:** V249.1      **文献标志码:** A      **文章编号:** 1671-637X(2016)05-0006-05

## UAV Target Tracking and Locating Based on Binocular Vision

WANG Ting-ting<sup>1</sup>, CAI Zhi-hao<sup>1,2</sup>, WANG Ying-xun<sup>1,2</sup>

(1. Beihang University, Beijing 100191, China; 2. Science and Technology on Aircraft Control Laboratory, Beijing 100191, China)

**Abstract:** Small-size Unmanned Aerial Vehicles (UAVs) have found an increasing application in GPS-denied environment. Stereo vision can be used for UAV target detecting, tracking and locating in complex environments. The thesis presents the techniques of binocular camera calibration, feature point detection, extraction and matching, binocular measurement and so on. It is implemented to track the specified target in image while realizing target positioning, with final calibration error within 0.2 pixels and positioning error within 10 cm. We made verification to the algorithm by taking a ground robot as the target to be tracked, and found that it can achieve a fine tracking result when the target has short disappearance time.

**Key words:** UAV; binocular vision; target tracking; camera calibration; CamShift algorithm

### 0 引言

近年来小型无人机向自主化、智能化的方向迅速发展,应用范围越来越广,尤其是在无卫星导航条件下的使用需求日渐强烈。目标监控是无人机的重要应用之一,采用立体视觉方法可实现无人机在复杂环境中的目标跟踪与定位。通过搭建双目视觉平台,采用适当的目标跟踪算法可以有效地获取目标的距离和位置,目标的运动方向、轨迹,大小等信息。双目视觉技术利用在不同视角下获得的感知图像,根据三角测量原理计算图像像素间的位置偏差,进而获取场景的三维信息<sup>[1]</sup>。文献[2]利用TLD目标跟踪算法实现了对单运动目标的较好跟踪;文献[3]利用单目摄像头实

现了小型无人机对地面目标的自动识别跟踪,并给出了控制策略;文献[4]利用双目视觉实现了六旋翼对目标的跟踪。目前,目标追踪问题主要表现为如何在复杂背景、光照变化、类似颜色物体、遮挡物等因素存在的情况下抓住目标的特点,进行有效跟踪,并保证实时性。鉴于以上内容,本文利用双目立体视觉进行目标追踪的研究具有重要的意义。

### 1 双目视觉系统标定

双目摄像机的准确标定是计算机视觉准确实现的基础和前提。目前摄像机标定方法有很多种,主要可分为摄像机传统标定方法和自标定方法。本文采用具有较好的鲁棒性和实用性的张氏标定法<sup>[5]</sup>进行摄像机标定。鉴于普通的COMS网络摄像头感光单元逻辑间距或者感光芯片的尺寸一般不得知,在标定结果判定上难度较大,因此,主要通过标定结果和理想参数的对比分析来判断标定结果的合理性。

收稿日期:2015-05-18      修回日期:2016-01-26

基金项目:航空科学基金(20135851043)

作者简介:王亭亭(1992—),女,山东济南人,硕士生,研究方向为导航制导与控制技术。

### 1.1 双目标定的原理

双目立体视觉系统可通过左右或者上下两个摄像头排列放置得到,通过对同一物体的观察,获得不同视角下的图像<sup>[1,6]</sup>。将摄像机模型简化为最简单的针孔模型,双摄像机成像模型如图1所示。

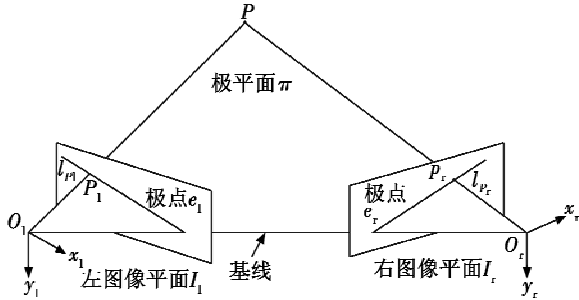


图1 双摄像机成像模型

Fig.1 Binocular camera imaging model

图中: $P$ 为空间一点; $O_l, O_r$ 分别为左右两个摄像头的光心; $I_l, I_r$ 为左右摄像机的成像平面; $P_l, P_r$ 分别为 $P$ 在左右两个摄像机平面上的投影点。

双目标定的结果是得到摄像机的内外参数,以及两个摄像机之间的关系。在标定过程中主要用到图像坐标系、摄像机坐标系和世界坐标系3种坐标系。图像坐标系中像素单位 $(u, v)$ 和物理单位 $(x, y)$ 两种表示方法的关系为

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{k} & 0 & u_0 \\ 0 & \frac{1}{l} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

式中,像素点大小为 $k * l$ ,单位为mm。令摄像机坐标系 $(X_0, Y_0, Z_0)$ 以摄像机镜头的光心为原点, $X_0, Y_0$ 轴与图像坐标 $x, y$ 轴平行, $Z_0$ 轴垂直于图像平面,则图像坐标系与摄像机坐标系之间的关系为

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} \frac{f}{k} & -\frac{f}{k} \cot \theta & u_0 \\ 0 & \frac{f}{l \sin \theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (2)$$

式中: $f$ 为摄像机焦距; $\theta$ 为摄像机坐标系的偏斜度,视为 $90^\circ$ 。世界坐标系可选空间任一点建立三维直角坐标,摄像机坐标系与世界坐标系的关系为

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{3 \times 3} & \mathbf{T}_{3 \times 1} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3)$$

式中: $\mathbf{R}$ 为旋转矩阵; $\mathbf{T}$ 为平移向量。

### 1.2 标定优化与结果分析

图2所示为采用2个LifeCam网络摄像头,摄像头

帧速率30帧/s,分辨率 $640 \times 480$ ,水平视角 $50^\circ$ ,垂直视角 $40^\circ$ ,摄像头光心相距约10cm,搭建双目平台。以一张 $8 \times 7$ 的棋盘格作为标定物,定标板每个棋盘格大小为 $25 \text{ mm} \times 25 \text{ mm}$ ,试验选取了36幅有效图像。

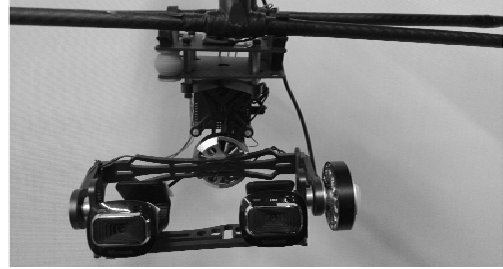
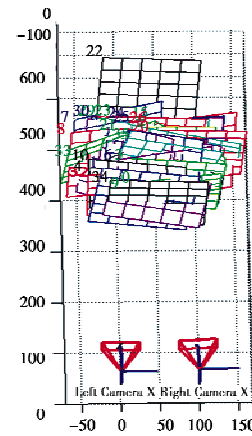


图2 双目视觉平台

Fig.2 Binocular vision platform

图3显示了图像相对于摄像头的位置关系。首先通过单目标定得到左右两个摄像机的内参矩阵,然后通过立体标定,得到右摄像机相对于左摄像机的旋转矩阵 $\mathbf{R}_r$ 和平移向量 $\mathbf{T}_r$ 。



a 双目下图像与相机的位置关系

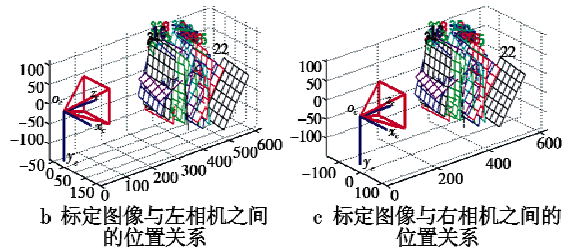


图3 标定图像与摄像机之间的位置关系  
(以摄像机为中心)

Fig.3 The relationship between the image and the camera (camera-centric)

根据双目视觉平台中摄像机与飞机机体的位置关系,为简化计算,在试验中使世界坐标系与左摄像机坐标系重合,则左摄像机相对于世界坐标系的旋转矩阵 $\mathbf{R}_l$ 为 $3 \times 3$ 单位阵,平移向量 $\mathbf{T}_l$ 为 $[0 \ 0 \ 0]^T$ ;右摄像机相对于世界坐标系的旋转矩阵为 $\mathbf{R}_r$ ,平移向量为 $\mathbf{T}_r$ 。

通过对80幅试验图标定后的误差进行分析,得出如果图像偏转过大或过小,或者图像放置较远或过偏,

标定误差会很大,对此应保证标定板偏转角在  $20^\circ \sim 70^\circ$  之间,且成像至少大概占图像的三分之一并尽量靠中间。在实际选取图像时,可以分别分析其误差图,对误差过大的图像改变窗口大小,重新标定,对于始终有较大误差的图像可以剔除,以此实现优化。图4所示为单个相机的重投影误差分布图,根据  $e = \sqrt{T_{\text{err}}/T_{\text{points}}}$  得平均误差在 0.2 个像素点以内。其中,  $T_{\text{err}}$  为总误差数,  $T_{\text{points}}$  为总点数。

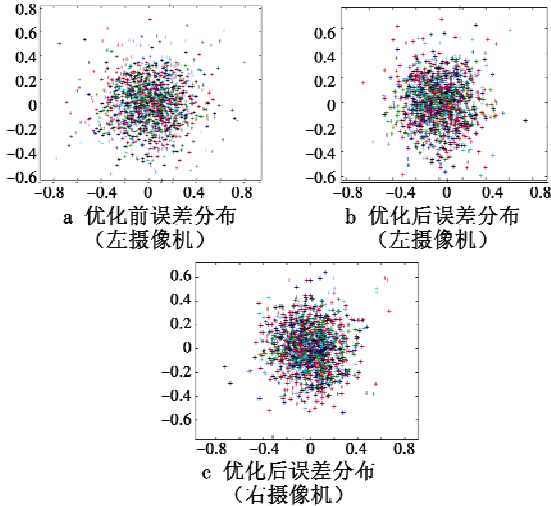


图4 误差分布情况

Fig.4 Error distribution

标定的结果如下。

左摄像头内参矩阵为

$$A_l = \begin{bmatrix} 673.83521 & 0 & 314.67757 \\ 0 & 672.90162 & 239.46904 \\ 0 & 0 & 1 \end{bmatrix};$$

右摄像头内参矩阵为

$$A_r = \begin{bmatrix} 672.2986 & 0 & 324.06032 \\ 0 & 671.80694 & 244.42814 \\ 0 & 0 & 1 \end{bmatrix};$$

左摄像头畸变系数为

$$k_l = [0.09506 \quad -0.54295 \quad 0.00198 \quad -0.00143 \quad 0.00000];$$

右摄像头畸变系数为

$$k_r = [0.09586 \quad -0.59101 \quad 0.00457 \quad -0.00272 \quad 0.00000];$$

右摄像头相对于左摄像头的旋转矩阵为

$$R_r = \begin{bmatrix} 0.9986 & 0.0028 & 0.0520 \\ -0.0017 & 0.9998 & -0.0217 \\ -0.0521 & 0.0216 & 0.9984 \end{bmatrix};$$

右摄像头相对于左摄像头的平移向量

$$T_r = [-99.90813 \quad 1.74141 \quad 3.82497]^T。$$

内参矩阵所得光心点坐标在(320 m, 240 m)附近,右摄像头相对于左摄像头的旋转矩阵接近单位阵,基本符合预期标定,从平移向量第一个参数可以看出,左

右摄像头之间的中心距为  $-99.90813$  mm,基本符合标定结果。

## 2 特征提取与匹配

特征点匹配是立体视觉第二大任务,意在建立两幅图像之间点到点的匹配关系。立体匹配的实质是求最优解,目前比较常用的匹配算法有基于特征的立体匹配算法,基于全局的立体匹配算法和基于局部的立体匹配算法<sup>[7-8]</sup>。为了保证算法实时性,采用匹配速度较快的基于特征的立体匹配算法,并采用 SURF 算法实现特征点提取。

SURF<sup>[9]</sup>特征提取算法是在 SIFT<sup>[10]</sup>算法的基础上提出的,在继承了 SIFT 算法鲁棒性的同时,提高了精度且匹配速度快,有很强的实时性。SURF 算法主要步骤为:建立积分图像→建立尺度空间<sup>[11]</sup>→快速 Hessian 极值点检测→Haar 小波建立描述算子→通过距离进行匹配<sup>[12]</sup>。文献[13]在进行特征点匹配误差分析时,将视差的计算结果分成独立的块,定量测评每块对匹配精度的影响。目前立体匹配算法大部分以此作为衡量算法优劣的标准。

利用 SURF 算法提取感兴趣区域的目标,通过匹配可以得到几组分别来自左右两幅图像的坐标点,每组点的组成为  $(u_1, v_1, u_2, v_2)$ ,其中感兴趣区域由目标追踪部分给出,以降低时间开销。根据前面的讨论,已知图像点坐标与世界点坐标的关系为

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_9 & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}。 (4)$$

左右图像的摄像机内参和外参已经获得,假设两个匹配点在图像坐标系中的坐标分别为  $(u_1, v_1)$  和  $(u_2, v_2)$ ,从而得到  $(X, Y, Z)$  相对  $(u_1, v_1)$  和  $(u_2, v_2)$  的关系为

$$Z = \frac{65896 + 4u_2}{49.440198 + 0.9706101u_1 + 0.01337458v_1 - u_2} (5)$$

$$X = \frac{u_1 - 314.6023}{673.803} Z (6)$$

$$Y = \frac{v_1 - 239.5313}{672.918} Z (7)$$

采用 Kalman 滤波对所得位置数据进行处理,进而得到目标相对飞机的位置。

## 3 目标追踪

目前常用的目标追踪算法有光流法、MeanShift 算

法、CamShift 算法、Kalman 滤波和粒子滤波等。每种算法均有自己的优势和劣势,需要根据实际应用采用合适的算法,或者对不同算法进行结合。比如将粒子滤波和均值漂移法相结合,或借助 Kalman 滤波算法比较强的鲁棒性,通过 Kalman 滤波算法估计物体的运动信息,以此解决遮挡和类似颜色物体存在造成的跟踪问题。如果物体运动随机性较大,则 Kalman 滤波算法会失效。因此,还需要结合其他算法或者研究新的追踪算法,同时还要保证实时性的要求<sup>[14]</sup>。

本文采用改进 CamShift 算法进行目标跟踪。基本的 CamShift 算法流程如图 5 所示。

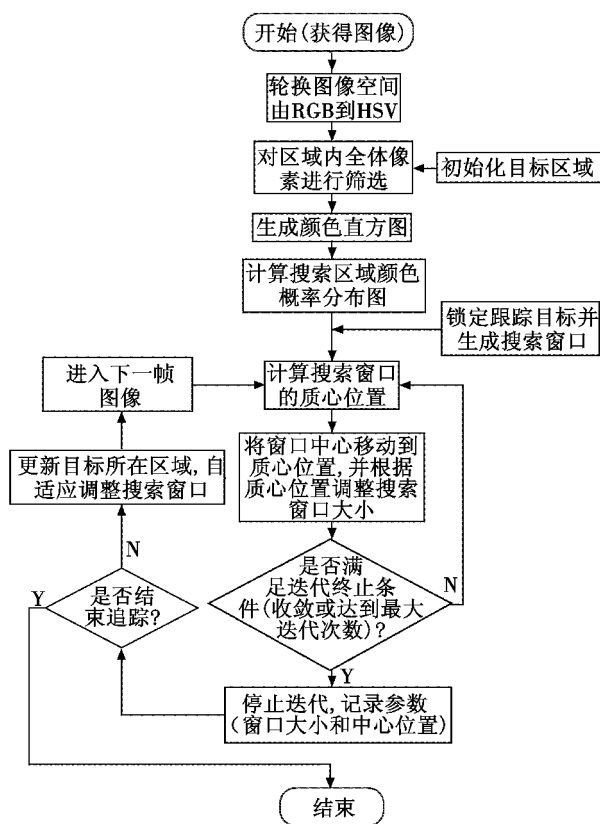


图5 CamShift 算法流程  
Fig.5 CamShift algorithm

在此基础上主要进行了如下改进。

1) 图像的预处理和颜色阈值的选取。

试验所用网络摄像头自带真彩处理功能,且右摄像头拍摄的图像对比度比左摄像头的高。因此,提高左摄像头图像的对比度并分析目标在两幅图像的 HSV 下 H 分量的直方图,选取最大值作为阈值。

2) 运动的预测降低相似目标干扰。

假设目标连续 3 次位置变化中方向和运动速度均不变,用近 2 次的目标位置信息可估计下一次目标可能的位置,在一定程度上可滤除视野中与目标颜色一致的静止干扰物。

至此,本文的 3 大环节已基本实现。图 6 所示为

整个算法的实现示意图。其中,虚线框内的双目标定部分为离线实现,其余均为在线实现。

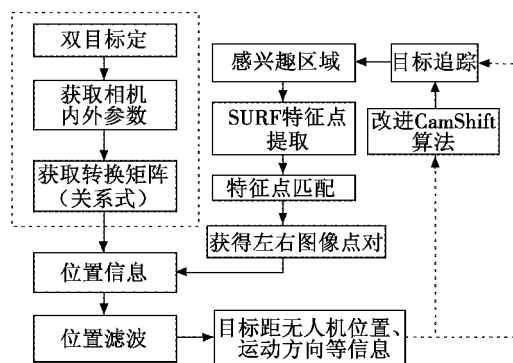


图6 3大环节关系示意图

Fig.6 Relationship between the three links

### 4 结果分析

如图 7 所示,采用 MTI-28A53G352,STM32(Cortex M3),2 轴无刷电机云台和双目视觉平台搭建基于双目视觉平台的目标追踪系统,IMU 的输出可以控制云台转动。

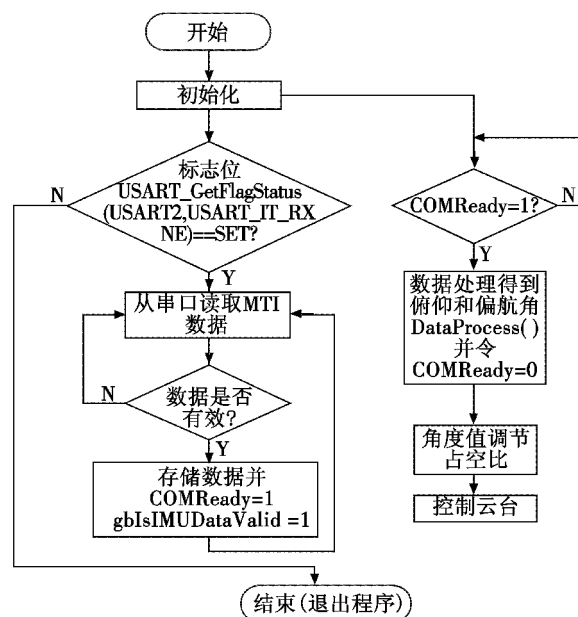


图7 MTI 控制云台

Fig.7 Control PTZ by MTI

如图 8、图 9 所示,在室内环境下,地面目标以 0.33 m/s 的速度做定速无规则运动,在摄像头固定和运动两种情况下做试验。

图 8 实际测量地面机器人中心点与摄像头的距离约 86 cm,机器人直径约 30 cm。保持地面机器人不动,摄像头在原位置做近似俯仰运动,测得的机器人位置信息如表 1 所示。将解算的位置信息与实际测量数据对比,平均误差约 10 cm。



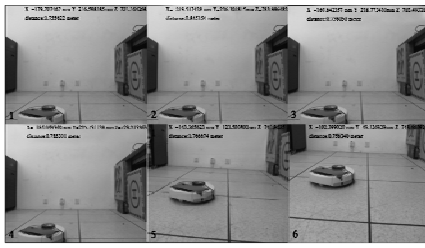


图 8 对追踪目标的位置信息显示(地面机器人)

Fig. 8 Location information for tracking (ground robot)

表 1 以地面机器人为目标位置信息

Table 1 Location information taking ground robot as target

	X/mm	Y/mm	Z/mm	相对距离/m
1	-179.70746	216.58818	701.16426	0.755622
2	-219.51724	226.30181	783.98648	0.845154
3	-169.84235	218.77341	703.49320	0.756050
4	-184.04934	225.73113	729.21336	0.785301
5	-143.36562	123.5038	741.95351	0.766674
6	-102.39902	49.13693	748.66092	0.758349

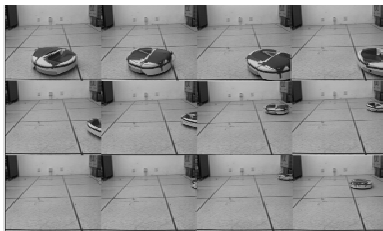


图 9 目标追踪(以地面机器人作为追踪目标)

Fig. 9 Target tracking taking ground robot as target

图 9 是对地面机器人的追踪。从左到右、从上到下分别选取了第 9 帧、第 13 帧、第 25 帧、第 33 帧、第 40 帧、第 54 帧、第 70 帧、第 75 帧、第 80 帧、第 84 帧、第 108 帧、第 135 帧的左摄像机拍到的图像,追踪效果良好,而且即便是地面机器人从视野中消失,只要能重回视野,依然能追踪得上。当摄像头与目标相对速度低于 1.5 m/s,距离在 3 m 以内时,跟踪效果良好。

## 5 结论

本文将双目视觉和目标追踪相结合,在对特定目标追踪的同时,能够得到目标的位置信息。当小型无人机与目标相对速度小于 1.2 m/s,距离小于 3 m 时,跟踪效果良好,满足小型无人机在室内环境下对双目视觉系统的要求,同时如果在该方案基础上进行相应改动可以实现低于 3 m 的飞行高度测量以及避障,这对小型无人机在无 GPS 的室内环境的应用具有很大的意义。

## 参考文献

[1] 陈胜勇,刘盛.基于 OpenCV 的计算机视觉技术实现[M].北京:科学出版社,2008.(CHEN S Y, LIU S.

Computer vision technology based on OpenCV[M]. Beijing: Science Press, 2008.)

- [2] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 34(7):1409-1422.
- [3] QUINTERO S A P, HESPANHA J P. Vision-based target tracking with a small UAV: optimization-based control strategies[J]. Control Engineering Practice, 2014(32):28-42.
- [4] 张梁,徐锦法.基于双目视觉的无人飞行器目标跟踪与定位[J].计算机工程与应用,2014,50(24):27-31.(ZHANG L, XU J F. Target tracking and locating for UAV based on binocular stereo vision[J]. Computer Engineering and Applications, 2014, 50(24):27-31.)
- [5] ZHANG Z Y. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11):1330-1334.
- [6] BAGGIO D I, EMAMI S, ESCRIVA D M, et al. Mastering OpenCV with practical computer vision projects[M]. Birmingham: Packt Publishing Ltd., 2012.
- [7] 丁菁汀.立体视觉在实际应用中的若干问题研究[D].杭州:浙江大学,2012.(DING J T. Research on several problems for the practical use of stereo vision[D]. Hangzhou: Zhejiang University, 2012.)
- [8] 曹之乐,严中红,王洪.双目立体视觉匹配技术综述[J].重庆理工大学学报:自然科学版,2015,29(2):70-75.(CAO Z L, YAN Z H, WANG H. Summary of binocular stereo vision matching technology[J]. Journal of Chongqing University of Technology: Natural Science Edition, 2015, 29(2):70-75.)
- [9] CHRISTOPHER E. Notes on the OpenSURF library[D]. Bristol: University of Bristol, 2009.
- [10] LOWE D G. Distinctive image features from scale-invariant key points[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [11] LINDBERGER T. Scale-space theory: a basic tool for analyzing structures at different scales[J]. Journal of Applied Statistics, 1994, 21(1/2):225-270.
- [12] 宋宁.基于 SURF 的运动目标检测与跟踪方法研究[D].昆明:昆明理工大学,2012.(SONG N. Study on the method of moving object detection and tracking based on SURF[D]. Kunming: Kunming University of Science and Technology, 2012.)
- [13] 罗桂娥.双目立体视觉深度感知与三维重建若干问题研究[D].长沙:中南大学,2012.(LUO G E. Some issues of depth perception and three dimension reconstruction from binocular stereo vision[D]. Changsha: Central South University, 2012.)
- [14] YILMAZ A, JAVED O, SHAH M. Object tracking: a survey[J]. Acm Computing Surveys, 2006, 38(4):81-93.