

单相机变焦图像深度估计技术研究

高宏伟^{1,2} 王慧科¹ 刘传银¹ 李 斌²

(沈阳理工大学信息科学与工程学院 沈阳 110159)¹

(中国科学院沈阳自动化研究所机器人学国家重点实验室 沈阳 110016)²

摘 要 对单相机变焦图像深度估计的若干关键技术进行了详细的研究。首先,从变焦相机的自身特点出发,研究了变焦相机的参数标定方法;其次,研究了基于 SIFT 的特征点匹配算法,并结合变焦图像的几何约束实现了特征点的鲁棒匹配;最后,建立了变焦图像的三维重建模型。基于真实采集图像的实验结果证明了相关算法的有效性和实用性。

关键词 变焦图像,标定,SIFT 匹配,深度估计

中图法分类号 TP249 文献标识码 A

Depth Estimation of Single Camera Zooming Image

GAO Hong-wei^{1,2} WANG Hui-ke¹ LIU Chuan-yin¹ LI Bin²

(School of Information Science & Engineering, Shenyang Ligong University, Shenyang 110159, China)¹

(State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China)²

Abstract Some key technologies for the depth estimation of single camera zooming image were investigated in detail. Firstly, camera parameters calibration was executed combining the characteristics of zooming camera. Then, robust feature matching based on SIFT was realized by geometry constraint of zooming image. Finally, the 3D reconstruction model of zooming image was established. The experiment results based on real sample images prove the validity and practicability of the related algorithms.

Keywords Zooming image, Calibration, SIFT matching, Depth estimation

1 引言

基于图像的深度估计是计算机视觉研究中的一个基本问题,它是迈向图像理解这个目标的关键一步,在移动机器人的环境感知中有着重要应用。通常用于测距的视觉系统大致可分为单目立体视觉、双目立体视觉和全向立体视觉。对于移动机器人而言,单目立体视觉系统由于具有性价比高、实时性好等优点,而具有广泛的应用前景^[1,2]。用单个相机实现测距任务大致可以分为两种方法,一种是被测物体不动,相机沿直线前后移动已知的距离,从而建模计算出物体距离相机的距离;另一种是采用电动或手动变焦镜头在两个焦距上分别成像,在两幅图像上匹配出同一个物体,从而测出距离。后者实现起来比较简单,所以研究较多。Ma 和 Olsen^[3]最早提出了利用变焦镜头实现深度估计的方法,在理论上表明了变焦图像可以提供关于深度的信息。Lavest 等^[4,5]对变焦镜头的光学特性做了精确的研究,提出精确的深度估计必须以厚透镜模型来描述变焦镜头。Asada 和 Baba 等^[6,7]根据实际镜头的结构提出了变焦、聚焦、光圈的三参数变焦镜头模型。Fayman 等^[8]将变焦图像的深度估计应用到了视觉跟踪领域,提

出了一种变焦跟踪的主动视觉技术,拓宽了变焦图像深度估计技术的应用范围。本文在全方位移动机器人实验平台上构建了一个变焦测距系统,对变焦相机的参数标定、基于 SIFT 的特征匹配和特征点的三维重建技术进行了详细的研究。基于真实采集图像的实验结果证明了相关算法的有效性和实用性。

2 变焦相机的标定

2.1 变焦图像的采集

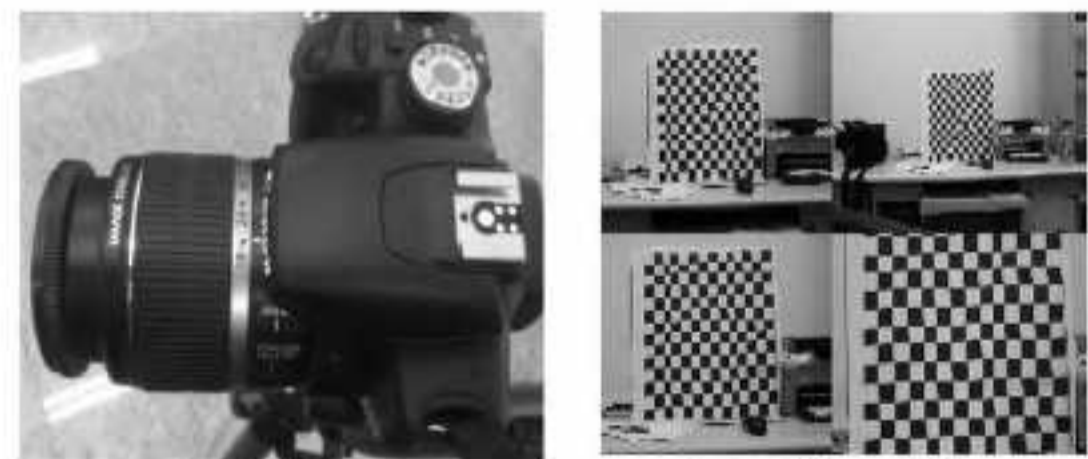


图 1 变焦图像的采集

对于变焦相机的标定而言,在一个固定的变焦镜头参数下,与普通的相机标定没有太大区别。但要建立描述变焦相

本文受中国博士后科学基金(2013M530954),中科院沈阳自动化研究所机器人学国家重点实验室开放课题(2012017)资助。

高宏伟(1978—),男,博士,副教授,主要研究方向为计算机视觉、智能控制,E-mail:ghw1978@sohu.com;王慧科(1988—),男,硕士生,主要研究方向为计算机视觉;刘传银(1989—),男,硕士生,主要研究方向为多无人系统协调控制;李斌(1962—),男,研究员,主要研究方向为特种机器人。

机的精确模型,通常需要建立一张表格,记录一系列相机在其变焦镜头不同设置下的标定结果。先在一个焦距下聚焦,然后获取不同焦距下的图像,可见两幅图像都很清晰,可以保证后续的图像匹配和深度重建的精度。变焦图像的采集如图 1 所示。

2.2 变焦图像的采集

由于只需要标定出两个焦距下的参数就可以完成深度估计,因此,按照前面介绍的设备和图像采集的方法,只标定变焦相机最大焦距 55mm(简写 F55)和最小焦距 18mm(简写 F18)下的内外参数即可。

图像畸变校正的方法如下:设空间点在相机参考坐标系下的坐标为 (X_c, Y_c, Z_c) ,在不考虑内部参数及畸变影响的情况下,该点投影到图像平面上的坐标为:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \end{bmatrix} \quad (1)$$

令 $r^2 = x^2 + y^2$,则引入畸变因子后的图像投影坐标为:

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = (1 + kc(1)r^2 + kc(2)r^4 + kc(5)r^6) \begin{bmatrix} x \\ y \end{bmatrix} + dx \quad (2)$$

其中, dx 为切向畸变,表示如下:

$$dx = \begin{bmatrix} 2kc(3)xy + kc(4)(r^2 + 2x^2) \\ kc(3)(r^2 + 2y^2) + 2kc(4)xy \end{bmatrix} \quad (3)$$

最后引入相机的内部参数,则空间点 (X_c, Y_c, Z_c) 在图像上的投影坐标为:

$$\begin{cases} x_p = fc(1)(x_d(1) + \alpha \cdot x_d(2)) + cc(1) \\ y_p = fc(2)x_d(2) + cc(2) \end{cases} \quad (4)$$

3 变焦相机的标定

利用 SIFT 算法进行图像匹配时^[9],匹配结果中不可避免地存在一些错误的匹配。而 SIFT 算法的应用和改进是否成功在很大程度上依赖于匹配水平的高低。因此如何判定匹配结果的正确与错误是首先要解决的问题。在变焦图像深度估计的研究中,有一条基本假设,即:理想状态下,以缩放中心为原点的图像坐标系中,匹配点的径向斜率是相同的(见图 2 中的匹配点 P_1 和 P_2)。显然,径向斜率相同是正确匹配点的必要条件。因此可以利用这个条件剔除匹配结果中的错误匹配点。

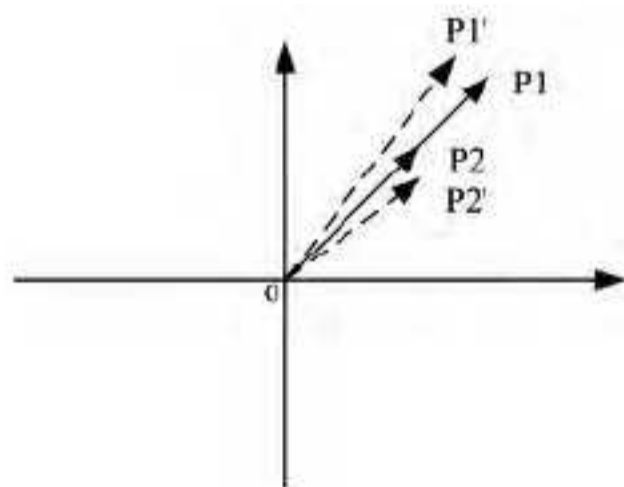


图 2 变焦图像中的理想匹配点和实际匹配点

但实际上,由于成像畸变等因素的影响,即使正确的匹配点,径向斜率也不可能完全相同(见图 2 中的匹配点 P_1' 和 P_2')。因此需要给出一个合理的容许度来筛选比较理想的匹配点,尽量剔除那些因为匹配算法所造成的错误匹配。但是不能随意给定可允许误差,最好通过实验给出。

假定在一对变焦图像中获得了 N 对匹配点,因为一对匹配点可以确定一条直线,所以可以获得 N 对匹配点的直线方程,设缩放中心为 (Z_x, Z_y) ,理想状态下缩放中心在每一条直

线上,则有矩阵方程:

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \dots & \dots \\ a_N & b_N \end{bmatrix} \begin{bmatrix} Z_x \\ Z_y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_N \end{bmatrix} \quad (5)$$

矩阵方程(5)可简写为:

$$A\theta = b \quad (6)$$

显然缩放中心 $\theta = (Z_x, Z_y)$ 可转化为最小二乘参数估计问题。即当 $N > 2$ 时,缩放中心由式(7)唯一确定:

$$\hat{\theta}_{LS} = (A^T A)^{-1} A^T b \quad (7)$$

缩放中心确定后,就可以通过一对匹配点和缩放中心确定的 2 条直线方程来计算匹配点径向夹角。设一对匹配点的坐标分别为 $(X_i, Y_i), (X_i', Y_i')$,则以缩放中心为顶点的任意一对匹配点径向夹角为:

$$\alpha_i = \arccos \frac{\vec{a}_i \cdot \vec{b}_i}{|\vec{a}_i| |\vec{b}_i|} \quad (8)$$

在匹配实际的图像之前,可以通过简单的结构化场景探测出匹配点径向夹角的大小水平和分布情况,从而为误匹配的去提供参考依据。假定探测出匹配点径向夹角分布区间为 $[a, b]$,利用传统的 SIFT 方法进行变焦图像的匹配时获得匹配点的集合为 $P = \{p_1, p_2, \dots, p_n\}$,而通过式(8)可获得匹配点径向夹角的集合为:

$$A = \{\alpha_i | \alpha_i = f(p_i), i \in [1, n]\} \quad (9)$$

在实际的误匹配去除时,一般情况下令 $a = 0$,而 b 等于最大可接受的匹配点径向夹角。

归纳起来,基于几何约束的变焦图像 SIFT 特征误匹配消除算法由下列步骤组成。

- Step 1 利用传统的 SIFT 方法获得变焦图像的匹配点的集合 P ;
- Step 2 根据匹配点的集合 P ,利用式(5)估计变焦图像的缩放中心;
- Step 3 利用式(6)计算匹配点径向夹角集合 A ;
- Step 4 利用式(7)获得理想匹配点的集合 P' 。

4 变焦图像的深度估计

针孔模型中,变焦等价于相机的光学中心沿光轴的移动(见图 3 中的 Δf)。不难发现,在针孔模型下物距(相机的光学中心和物体之间的距离)的变化量等价于焦距的变化量(光学中心的平移量)。在该模型中,利用至少两个不同焦距所成的图像便可以获得物体的深度信息,如图 3 所示,物体的深度计算公式为:

$$Depth = \frac{\Delta f r_1 f_2}{f_1 r_2 - f_2 r_1} \quad (10)$$

其中, r_1 和 r_2 是变焦图像中一对匹配点的径向半径,可以由前述 SIFT 匹配点对的图像坐标确定。

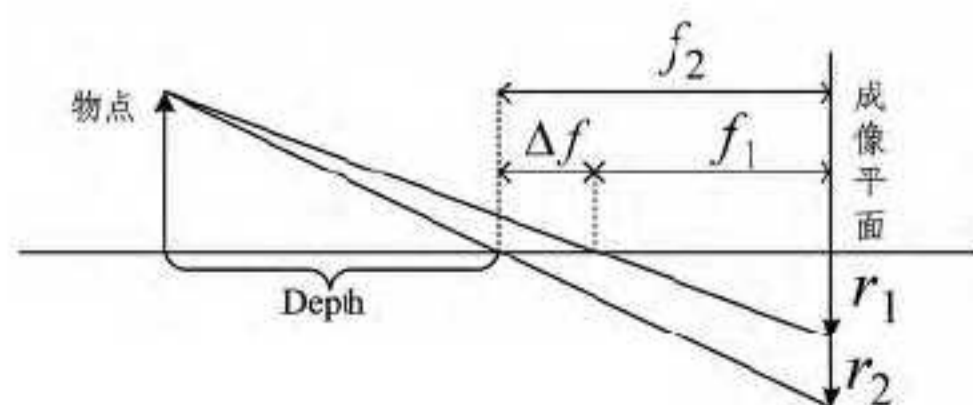


图 3 变焦深度估计的针孔模型

5 实验结果与分析

5.1 标定结果

实验使用平面模板的方格尺寸为 30 毫米。原始图像的像素分辨率为 2352×1568 。本文一共采集了 10 组标定的图像数据,每组数据都包含 F18 和 F55 两个焦距下采集的图像。每个焦距下一般采集 20 幅左右的图像,以下是最终的标定实验结果与分析。

1) 焦距的标定结果

图 4 所示为焦距 f 的标定结果,标定工具箱直接的数据是以像素为单位的。图 4(a) 是焦距 F18 标定结果的变化曲线,图 4(b) 是焦距 F55 标定结果的变化曲线。尽管采集的数据的时刻不同、聚焦不同,标定时图像的幅数也不同,但还是获得了较稳定的结果。

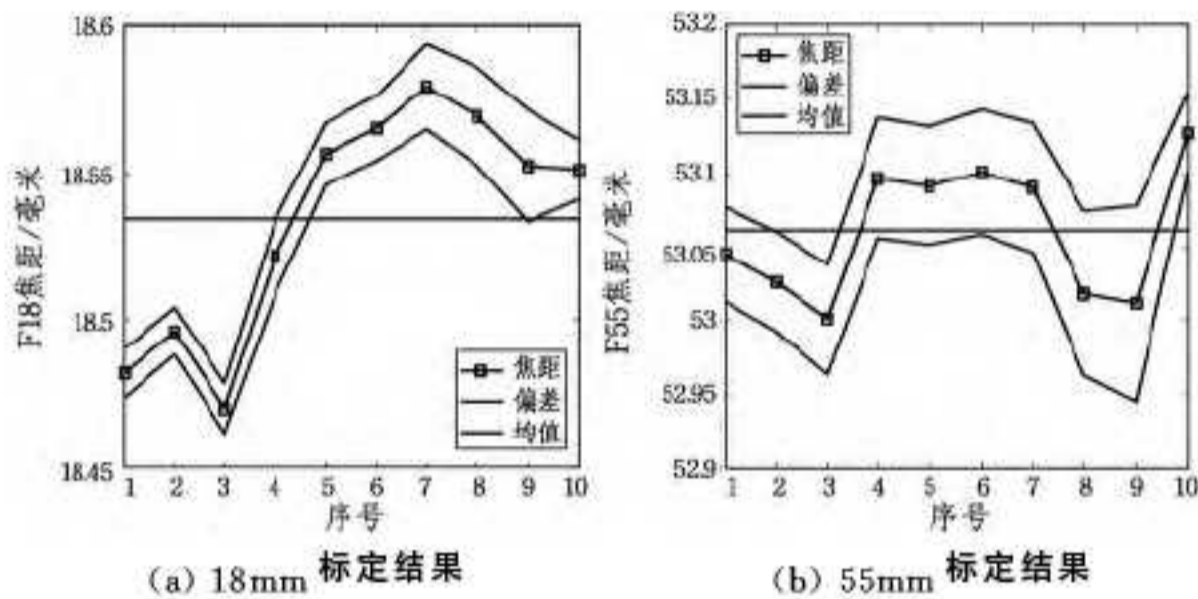


图 4 焦距 f 的标定结果

2) 畸变系数的标定结果

图 5 所示为畸变系数 k_c 的标定结果,畸变系数是无量纲数据。图 5(a) 是 F18 下畸变系数 k_c 标定结果的变化曲线,图 5(b) 是 F55 下畸变系数 k_c 标定结果的变化曲线。它们相对其它参数畸变系数的标定结果更加稳定,尤其是焦距 F18 下的畸变系数。但是不同焦距下变焦镜头的畸变系数 k_c 完全不一样。可见畸变系数随着焦距的变化而变化,但在某一焦距下是相对稳定的。

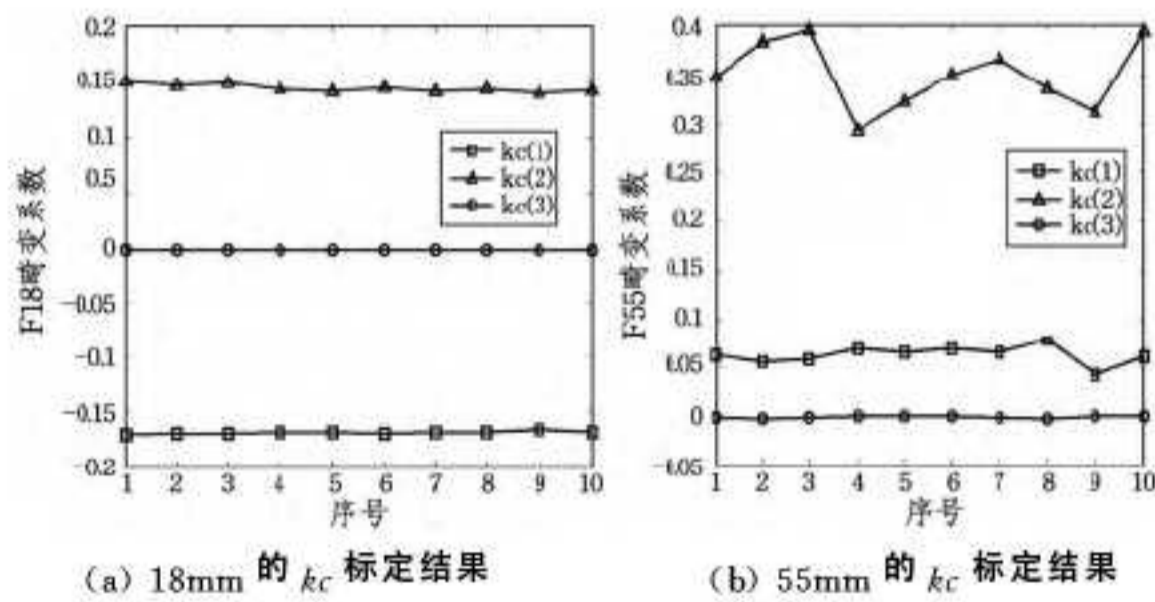


图 5 畸变系数 k_c 的标定结果

5.2 匹配与三维重建结果

利用上述算法对其它变焦图像进行了处理,依然取得了良好的效果,实验结果如图 6 所示。图 6(a) 是花的原始变焦图像对。图 6(b) 为 SIFT 算法直接匹配结果,匹配点总数 138 个,可以看出存在误匹配点对。图 6(c) 为去除误匹配的结果,剩余匹配点数 127 个,占总数的 92.02%,基本没有误匹配,可以看出利用变焦图像的几何约束有效地去除了误匹配点。图 6(d) 为三维重建结果。本文采用最小内角最大化原则的 Delaunay 三角剖分法进行三维数据的处理。这样,深度图的纵横坐标可以由场景点在相机坐标系下的真实坐标构成,从而实现物体表面的三维重建。

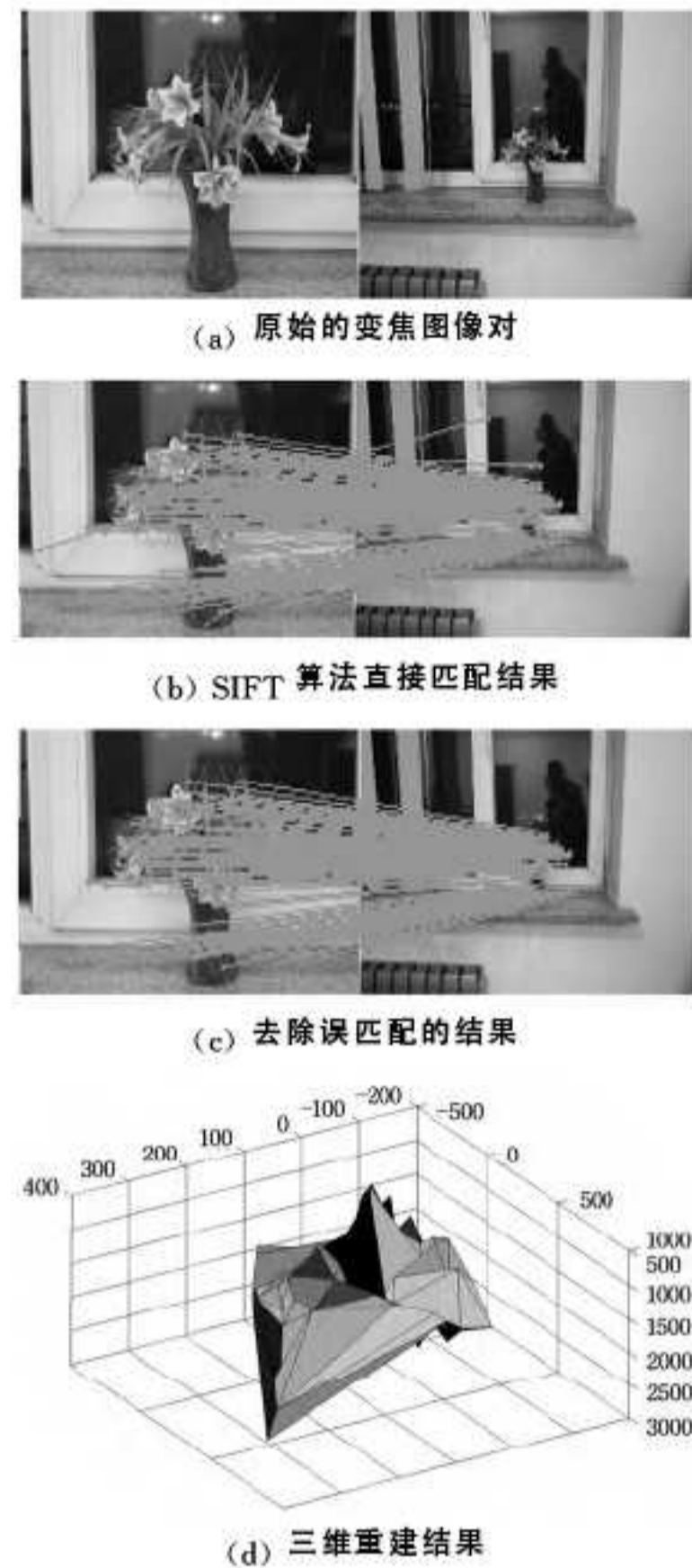


图 6 花的变焦图像对

结束语 为了实现变焦图像的深度估计,本文建立了一个面向移动机器人的单相机变焦测距系统,机器人只需在某个位置停留,并对相机进行自动变焦,采集两幅图像,即可实现特征点的深度测量。文中从变焦相机的自身特点出发,研究了变焦相机的标定、匹配与三维重建,进行了非结构化场景的深度估计实验,实验结果表明了相关算法的有效性和实用性。

参考文献

- [1] 庄严,陈东,韩建达,等. 移动机器人基于视觉室外自然场景理解的研究与进展[J]. 自动化学报,2010,36(1):1-11
- [2] 张学习,杨宜民,蔡述庭. 移动机器人前向单目视觉的建模研究[J]. 控制与决策,2012,27(5):792-796
- [3] Ma J, Olsen S I. Depth from Zooming[J]. Journal of the Optical Society of America, 1990, 7(10):1883-1890
- [4] Lavest J M, Rives G, Dhome M. Three-dimensional Reconstruction by Zooming[J]. IEEE Transactions on Robotics and Automation, 1993, 9(2):196-207
- [5] Lavest J M, Rives G, Dhome M. Modeling an Object of Revolution by zooming[J]. IEEE Transactions on Robotics and Automation, 1995, 11(2):267-271
- [6] Asada N, Baba M, Ode A. Depth from Blur by Zooming[C]// Proc. Vision Interface, 2001:165-172
- [7] Baba M, Asada N, Oda A, et al. A Thin Lens Based Camera Model for Depth Estimation from Defocus and Translation by Zooming[C]// Proceedings of ICVI 2002:274-281
- [8] Fayman J F, Sudarsky O, Rivlin E, et al. Zoom Tracking and Its Applications[J]. Machine Vision and Applications, 2001, 13(1):25-37
- [9] Ke Y, Sukthankar R. PCA-SIFT: A More Distinctive representation for Local Image Descriptors[J]. IEEE Conf. on Computer Vision and Pattern Recognition, 2004, 2(2):506-513