

基于双目视觉的三维测量技术研究

周科杰^{1,2}, 冯 常¹

(1. 中国科学院光电技术研究所, 成都 610209; 2. 中国科学院大学, 北京 100049)

摘要: 文章对基于双目视觉的三维测量技术研究进行了深入挖掘, 主要包括摄像机模型、双目相机的标定、图像校正、立体匹配等; 同时借鉴机器学习的分类思想, 将角点提取转化为二分类问题, 利用随机森林算法来实现角点提取, 再利用随机森林的预测结果来实现亚像素级角点提取; 该方法相对于传统三维测量中的角点提取算法具有更好的自动化性能, 能避免角点集群现象, 能实现较高精度的亚像素级角点提取, 获得更高精度的二维像素坐标; 从而利用高精度的二维像素坐标来获得点的三维世界坐标, 这也是基于双目视觉的三维测量技术的基础与核心。

关键词: 双目视觉; 三维测量; 角点提取; 随机森林; 立体匹配

Three-dimensional Measurement Technology Based on Binocular Vision

Zhou KeJie^{1,2}, Feng Chang²

(1. Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: In this paper, the research of 3D measurement technology based on binocular vision has been deeply explored, including camera model, binocular camera calibration, image correction, feature extraction, feature point matching and so on. At the same time, based on the classification idea of machine learning, the corner extraction is transformed into a binary classification problem, and the random forest algorithm is used to achieve corner extraction, and the sub pixel corner extraction is achieved by random forest predictions. Compared with the traditional 3D corner detection algorithm, this method has better automation performance, and can avoid the phenomenon of corner clusters and achieve high-precision sub-pixel corner extraction to obtain more precise 2D pixel coordinates. The high-precision 2D pixel coordinates are used to get the 3D world coordinates of points, this is also the basis and core of the 3D measurement technology based on binocular vision.

Keywords: binocular vision; 3D measurement; corner extraction; random forest; stereo matching

0 引言

近年来, 基于机器视觉的三维测量技术不断发展, 尤其是基于双目视觉的三维测量技术由于其非接触测量、自动化程度高等优点在机器人, 无人机, 工业检测等领域有着广泛的应用前景。全燕鸣等将基于双目视觉的三维测量技术应用于工件的在机测量中, 实现了工业环境中双目视觉三维测量技术的应用^[1]。朱先锋等在基于 .net 与 halcon 的软件平台上通过计算目标的外包矩阵从而实现目标的三维测量^[2]。田小超等结合神经网络算法提出了一种基于 BP 神经网络的双平面标定法^[3], 利用神经网络的非线性映射能力对摄像机模型进行优化, 从而提高三维测量精度。

基于双目视觉的三维测量技术就是利用双目相机拍摄的图像, 将目标在图像上的二维像素坐标转化为三维世界坐标。当恢复了目标上的必要关键点之后, 物体的整个外形和位置就可以被唯一确定。本文在对基于双目视觉的三维测量技术进行深入研究之后, 针对三维测量中传统角点提取算法需要手动调整参数、角点集群、精度较低等弱点提出了一种基于随机森林的亚像素级角点提取算法, 该算

法不但克服了传统算法的缺点, 同时具有很高的精度。

1 双目视觉系统

双目立体视觉是基于视差原理的视觉成像模型^[4], 利用视差重建目标的三维信息。如图 1 为基本的双目立体视觉原理图, 左右摄像机同时对目标点 $P(x, y, z)$ 成像, 并分别在各自的成像平面上留下 $P_l(x_l, y_l)$, $P_r(x_r, y_r)$ 点, 同时 O_lC_l , O_rC_r 相互平行, P_lP_r 满足极线约束, 视差 $D = x_l - x_r$, P 点在摄像机坐标系下的坐标为 (x_c, y_c, z_c) 。根据相似三角形原理, P 点在摄像机坐标系下的坐标可以表示为:

$$\begin{cases} x_c = b \frac{x_l}{x_l - x_r} = b \frac{x_l}{D} \\ y_c = b \frac{y_l}{x_l - x_r} = b \frac{y_l}{D} \\ z_c = b \frac{f}{x_l - x_r} = b \frac{f}{D} \end{cases} \quad (1)$$

即我们只要能知道目标点在左右成像平面上的成像平面坐标, 就可以求出目标点的三维坐标, 从而获得目标点的三维信息。

1.1 双目视觉成像模型

设相机模型中的四个平面坐标系分别为: 像素坐标系 (u, v) 、成像平面坐标系 (x, y) 、相机坐标系 (x_c, y_c) 、

收稿日期: 2018-06-02; 修回日期: 2018-07-31。

作者简介: 周科杰(1993-), 男, 四川巴中人, 硕士研究生, 主要从事图像处理、机器学习方向的研究。

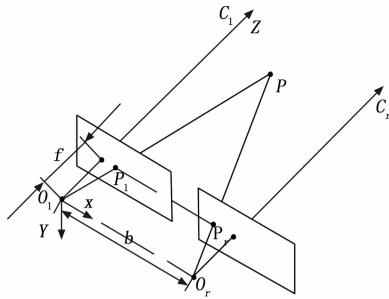


图 1 双目视觉原理图

世界坐标系 (x_w, y_w, z_w) ^[6]。双目视觉成像模型为式 (2), 建立了世界坐标系到像素坐标之间的转换关系。

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/d_x & r & u_0 \\ 0 & 1/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2)$$

1.2 双目视觉相机标定

根据式 (2) 可以看出, 双目立体视觉需要先确定相机的内外参数, 建立空间点到像素点的映射关系, 这个过程就是摄像机标定。张正友标定算法^[7]以其标定精度高、操作简单、鲁棒性好等特点成为相机标定领域最经典、最常用的标定算法之一, 其核心部分主要为:

- 1) 在不考虑摄像机畸变的条件下, 求取摄像机的内外参数;
- 2) 利用最小二乘法计算畸变系数;
- 3) 利用第 1, 2 步的结果再次优化内外参数, 使误差最小。

2 双目视觉立体匹配

立体匹配是双目立体视觉三维测量最核心, 也是最困难的部分。立体匹配算法主要分为基于区域的立体匹配算法、基于特征的立体匹配算法^[11]。其中基于特征点的立体匹配算法以其匹配效率高、精度高、抗噪声能力强等特点成为最有效的匹配算法之一。基于特征点匹配的立体匹配主要包括图像预处理、特征点提取、特征点匹配。其中预处理通常包括目标提取、极线校正两个部分。

待测目标提取:

待测目标的提取对于双目视觉的三维测量不是一个必需的步骤, 尤其是在待测目标工作环境较为简单的条件下。但如果目标所处环境较为复杂, 则先对目标进行预提取, 能极大的提高立体匹配算法的精度、优化立体匹配算法的时间复杂度和空间复杂度。目标提取最常用的算法包括形态学背景去除算法、帧差背景去除算法、图像分割算法等^[9]。

2.2 图像校正

在上文中双目视觉基本原理公式 (1) 时, 曾提到点 P_l, P_r 处于平行于基线的同一水平面上, 有 $y_l = y_r$, 这是

理想的平行光轴双目立体视觉模型。而现实中搭建双目立体视觉模型时, 或多或少都存在偏差。图像校正就是通过通过对左右图像进行几何变换使得两幅图像对应匹配点处于同一条直线上, 这称之为极线约束^[8], 如图 2 所示为极线校正前后的示意图。

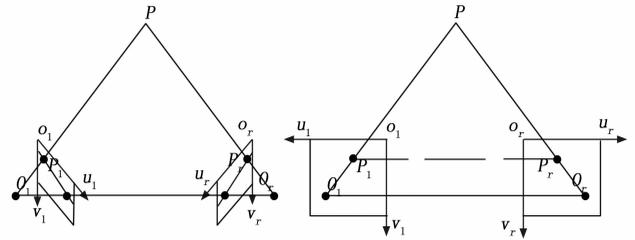


图 2 极线校正

极线约束使得双目立体视觉在进行立体匹配时, 大幅度减少匹配点搜索空间, 从而能够有效提高匹配的准确度, 减少匹配算法的时间复杂度和空间复杂度。

2.3 特征点提取

图像特征点能够反映图像本质特征, 能够标识图像中目标物体。最常用的特征点提取算法包括基于轮廓线的边缘检测和基于灰度值变化的角点检测^[1]。特征点提取是整个双目视觉三维测量中对测量精度影响非常高的一个环节。传统的如 Harris 角点提取算法主要有以下缺点:

- 1) 只能进行像素级角点提取;
- 2) 难以克服角点集群问题;
- 3) 智能化程度不够, 常常需要进行比较繁琐的阈值等参数的调节;
- 4) 引入了很多不够合理的先验, 如高斯分布、线性化等过程。

对此, 本文将特征点提取转化为机器学习的二分类问题, 即预测图像中的像素点是角点的概率, 提出了一种基于随机森林的亚像素角点提取算法。

2.3.1 随机森林

随机森林^[10]是机器学习中最知名, 最有效的算法之一, 它基于决策树和 bagging 的思想构建出强大的分类回归森林。随机森林具有结构简单, 计算复杂度小, 容易并行化且具有极强的泛化性能等优点。如图 3 为随机森林原理图, 其基模型为 CART 树, 通过对样本集的样本和特征的有放回采样来构建多颗 CART 树, 最后将每棵树的预测结果平均, 得到像素点为角点的概率。

结合随机森林的预测结果, 进行亚像素角点提取:

$$\text{corner} = \frac{\sum_{s=-1}^1 \sum_{t=-1}^1 P(x+s, y+t)^N C_{x+s, y+t}}{\sum_{s=-1}^1 \sum_{t=-1}^1 P(x+s, y+t)^N} \quad (3)$$

其中: $P(x+s, y+t)$ 为 $(x+s, y+t)$ 处为角点的概率, $C_{x+s, y+t}$ 为 $(x+s, y+t)$ 处的坐标, corner 为最终预测的亚像素角点坐标, $P(x, y)$ 为 (x, y) 所在 3×3 邻域内概率最大的像素点, N 为整数且。

2.3.2 特征向量提取

结合特征点自身的性质一与其邻近像素点灰度值高度

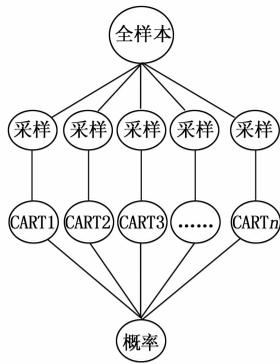


图 3 随机森林

相关,同时出于数据集大小、算法效率的考虑选用了 Haar 中的部分核来生成特征向量,如表 1 所示。

表 1 生成特征向量的核

核类型	核
两元素核	$[1 \ -1], \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
三元素核	$[1 \ -1 \ 1], \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$
四元素核	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
九元素核	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

2.3.3 基于随机森林的角点提取算法

基于随机森林角点提取算法模型训练过程如下。

- 1) 双目相机采集含有目标的图像 K 幅;
- 2) 对图像进行角点标注,需要确保角点标注的准确性;
- 3) 在每幅图像中除角点以外的位置进行采样,包括背景、物体边缘处等等,尽可能保证采样点含有不同的模式,不含有信息重复较多的点;最好保证正负样本的比例均衡;
- 4) 计算角点和采样的非角点的特征向量,窗口大小为 $N * N$;
- 5) 角点记为 1,非角点记为 0,生成训练数据;
- 6) 使用随机森林算法对模型进行充分训练;
- 7) 将训练好的模型保存下来;
- 8) 当双目相机采集到图像后,如果图像分辨率太高,且目标所占图像比例较小,需要对图像进行预处理,以降低整个角点提取的复杂度;
- 10) 对预处理后的点进行二分类预测,精确地找出所有角点。

2.4 特征点匹配

特征点匹配就是通过一定的匹配算法,将图像上已经

检测到的特征点同另一幅图像上的特征点进行匹配,在理想情况下这两点将对现实中的同一点。因此,特征点匹配本质上是一种相似度度量。

根据不同的相似度度量标准,有不同的特征点匹配算法。通常使用 MAD, SAD 算法, SSD 算法, NCC 算法四种相似度度量准则进行特征点匹配。这四种算法前三种本质上是一个原理,就是在计算两个窗口中对应像素点的灰度值差,从而衡量两个窗口的相似度。这三种算法对光照较为敏感,而基于双目视觉的立体匹配由于左右相机视角问题会导致左右相机采集的图像亮暗程度不一,因此这三种算法不适用于双目视觉的立体匹配中。而 NCC 算法则对灰度值进行了归一化处理,降低了光照对立体配的影响,对双目视觉来说较为适用。

3 实验

3.1 实验设备

本次实验所使用的实验设备如图 4 所示,主要包括 FUJIFILM 工业摄影机,高精度标定板,高精度量块以及便携式工控机。相关参数见表 2。



图 4 实验设备

表 2 实验设备主要参数

实验设备	参数	参数值
摄像机	分辨率(pix)	2448 * 2048
	焦距(mm)	25
	光圈	F1.4~F22
	最小物距(m)	0.1
	重量(g)	315
标定板	尺寸(mm)	100 * 100
	圆点阵列	7 * 7
	圆点直径(mm)	6.25
	圆点中心距(mm)	12.5
量块	精度(mm)	0.001
	长(mm)	60
	宽(mm)	35
	高(mm)	9

3.2 相机标定

根据 2.2 节所介绍的双目立体视觉标定流程,对相机进行标定,其中拍摄的标定板照片组数 $K=15$ 。如图 5 为拍摄的 15 组标定板图像。同时标定过程中,需要注意两

点: 一是保持标定板背景的单一, 最好是黑色或者白色; 其次要防止光照对相机标定的影响。最终我们求得相机的内外参数见表 3, 标定平均误差为 0.034 2 个像素。

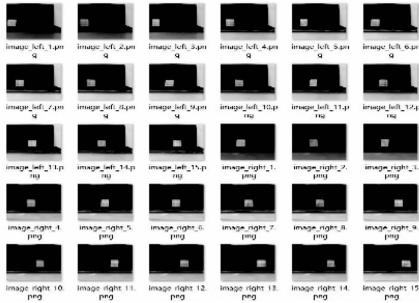


图 5 15 组标定板图像

表 3 相机内参数标定结果

参数类型	标定参数	标定结果	
		左相机标定结果	右相机标定结果
内参数	焦距 (mm)	0.0240	0.0240
	畸变系数 ($1/m^2$)	-180.5661	-139.5703
	单个像素宽 (μm)	3.45	3.45
	单个像素高 (μm)	3.45	3.45
	主点 x 坐标 (pix)	1202.7564	1192.6069
	主点 y 坐标 (pix)	999.7325	1002.0926
	图像宽 (pix)	2448	2448
	图像高 (pix)	2048	2048
外参数	旋转矩阵 $R(^{\circ})$	[0.1836, 0.0014, 0.0026]	
	平移矩阵 $T(m)$	[0.1133, 357.7735, 0.2940]	

3.3 三维测量实验

我们在角点提取算法中选用的随机森林算法是基于 Python 的著名机器学习库 scikit-learn 的 RandomForestClassifier 接口, 这是一个由前谷歌工程师 David Cournapeau 发起的开源机器学习项目, 包含了当今机器学习领域内绝大多数分类、回归、聚类、降维等算法。同时 scikit-learn 提供交叉验证, 这对于构建一个稳定的机器学习模型有极大的好处。三维测量主要试验步骤如下:

- 1) 双目相机采集含有量块的图像 90 幅 (左右图像各 45 幅), 每幅图像 6 个角点;
- 2) 对图像进行角点标注, 在每幅图像中除角点以外的位置进行采样, 包括背景、物体边缘非角点处, 角点总共为 540, 非角点 540;
- 3) 以被选择的点为窗口中心, 在 5×5 的窗口内计算点的特征向量;
- 4) 角点记为 1, 非角点记为 0, 生成训练数据;
- 5) 使用随机森林算法对模型进行充分训练, 同时进行 5 折交叉验证;
- 6) 将训练好的模型保存下来, 以便重复使用;
- 7) 对待预测的图像进行预处理, 剔除掉绝大多数平滑的点, 减少计算量;
- 8) 对预处理后的点进行二分类预测, 每个位置选取概

率值最高的点作为角点。

9) 在第 8 步预测的点周围 3×3 的窗口内, 利用式 3 计算亚像素角点坐标 ($N=3$)。

其中随机森林模型训练的主要参数设置如表 4, 仅供参考, 在训练模型时根据实际情况进行调节。

表 4 随机森林关键参数设置

参数	参数含义	参数作用	参数值
$n_estimators$	随机森林中 CART 树数量	保证模型泛化性能	10
$criterion$	CART 树划分指标	CART 树节点划分标准	'gini', 基尼系数
$max_features$	每颗 CART 树最大特征维度	保证模型泛化性能	'sqrt', 最大特征数量的均方根
max_depth	CART 树最大深度	保证模型精度	10
$min_samples_leaf$	叶子节点上最小样本数量	防止模型过拟合	7
$bootstrap$	有放回的采样	保证模型泛化性能	True
$min_impurity_split$	最小划分纯度	防止模型过拟合	$1e-7$
n_jobs	多核计算	加速模型收敛	4

3.4 误差分析

当像素点的三维世界坐标被还原之后, 一个孤立的点, 我们难以度量其准确性。因此, 我们计算出量块多个角点的世界坐标, 通过两点坐标计算长度来衡量三维世界坐标重建的准确性。我们使用不同的 5 组图像, 使用 3.2 中训练好的模型进行检测, 将 5 次的测量结果平均, 如表 5 所示。

表 5 测量结果

真实距离	随机森林像素级	随机森林亚像素级
9	9.091	9.064
35	34.912	35.067
60	59.873	59.895

通过表 5 可以看出, 使用随机森林来提取特征点, 进而三维测量, 是具有较高精确度的, 随机森林的亚像素级能进一步减小测量误差。

4 结语

本文深入研究了利用双目立体视觉进行三维测量的技术与方法, 同时针对其中的难点问题, 引入了基于随机森林的机器学习算法来进行特征点提取。基于机器学习强大的自学习能力, 该算法具有以下优点:

- 1) 传统算法如 harris 等特征点提取算法, 每次测量都需要手动设置参数和阈值; 而基于随机森林等机器学习的算法一旦模型训练好之后, 后续测量将是比较自动化的过程。
- 2) 传统的特征点提取算法几乎都存在角点集群问题,