

基于双目视觉和约束条件的行人目标定位

杨志光¹ 张永林¹ 谢经明² 陈幼平²

1. 武汉轻工大学, 武汉, 430023

2. 华中科技大学国家数控技术工程研究中心, 武汉, 430074

摘要:针对行人检测中计算量大、训练分类器耗时和无法满足实时性要求等问题,提出了一种基于双目视觉的行人目标定位方法。该方法利用图像处理技术获取候选轮廓,将轮廓的几何特征作为约束条件来筛选候选轮廓;利用双目视觉获取轮廓的深度信息后,将深度信息作为约束条件对候选轮廓进行进一步筛选。通过深度信息和几何信息的共同约束,识别出行人的头部轮廓,从而实现对行人目标的定位。实验结果表明,该方法减小了计算量,提高了检测精度。

关键词: 双目视觉;图像处理;行人检测;目标定位

中图分类号: TP216

DOI:10.3969/j.issn.1004-132X.2016.23.009

Target Positioning of Pedestrian Based on Binocular Vision and Constraints

Yang Zhiguang¹ Zhang Yonglin¹ Xie Jingming² Chen Youping²

1. Wuhan Polytechnic University, Wuhan, 430023

2. National NC System Engineering Research Center, Huazhong University of Science and Technology, Wuhan, 430074

Abstract: When confronting the problems in pedestrian detection such as large amount of calculation, time-consuming of classifier training and unfulfilled real-time requirements, a pedestrian detection method was proposed based on binocular vision. In this method, a lot of candidate contours might be obtained by processing image, and the geometrical characteristics of contours were used as a constraint to filter candidate contours. In the meantime, binocular camera might be used to get depth informations of the candidate contours, and the depth informations were used as a constraint to filter the candidate contours. By the constraints of depth informations and geometrical informations, contours of pedestrians' heads might be identified and the pedestrians' localization might get. The experimental results show that the method reduces the amount of calculation, and enhances the detection accuracy.

Key words: binocular vision; image processing; pedestrian detection; target positioning

0 引言

机器视觉利用摄像机代替人眼对物体进行检测、跟踪和识别^[1]。当前,获取实时客流量对地铁

基金项目:湖北省科技计划资助项目(2014BEC056)

部门提高工作效率有着重要的现实意义^[2]。因此,笔者将机器视觉运用于客流量统计系统,提出一种基于双目视觉的目标定位方法,实现对行人的检测。

收稿日期:2016-07-29

[11] 李冬伟,白鸿柏,杨建春,等.金属橡胶弹性元件实验建模研究[J].兵器材料科学与工程,2005,28(3):7-10.

Li Dongwei, Bai Hongbo, Yang Jianchun, et al. The Experiment Study of Metal Rubber Elastic Element Model[J]. Journal of Weapon Materials Science and Engineering, 2005,28(3):7-10.

[12] 刘家燕,程志峰,王平.机载光电吊舱橡胶减振器的设计与应用[J].中国机械工程,2014,25(10):1308-1312.

Liu Jiayan, Cheng Zhifeng, Wang Ping. The Design and Application of Airborne Electro-optical Pod Rubber Shock Absorber [J]. China Mechanical

Engineering, 2014,25(10):1308-1312.

[13] 李宇明,彭威,白鸿柏,等.金属橡胶材料宏观和细观力学模型[J].机械工程学报,2005,41(9):38-41. Li Yuming, Peng Wei, Bai Hongbo, et al. Mechanical Model of Metal Rubber Material in Macro and Micro Level[J]. Journal of Mechanical Engineering, 2005,41(9):38-41. (编辑 王旻玥)

作者简介:余慧杰,男,1978年生。上海理工大学机械工程学院讲师、博士。主要研究方向为电子设备结构优化设计、隔振系统设计及人机工程动态仿真。发表论文10篇。刘文慧,女,1989年生。上海理工大学机械工程学院硕士研究生。王亚苏,男,1989年生。上海理工大学机械工程学院硕士研究生。

传统行人检测技术易受到行人姿态、光照、背景、阴影、遮挡等因素的影响^[3],使得行人检测成为机器视觉领域的难点。目前,行人检测方法一般分为外观特征提取和分类器学习两种。外观特征提取算法又可分为标量特征运算和矢量特征运算。其中,标量特征运算的速度快、实时性好,但易受外部光线因素的影响,鲁棒性较差^[4];矢量特征运算主要考虑特征的梯度向量,鲁棒性较好,但计算量大,难以满足系统实时性的要求。常用的分类学习方法有支持向量机^[5]、神经网络^[6]和自适应增强算法^[7]。基于分类学习的检测方法具有较好的鲁棒性,正确率也较高,但分类器的训练比较耗时,且算法复杂度较高,难以满足实时监控的需求。

本方法的实际应用环境——地铁站的光照条件稳定,故以标量特征运算为基础,提出一种基于双目视觉的行人定位方法。通过双目视觉系统来获取物体相对于摄像机的深度信息,利用深度信息排除低于指定高度的目标,如行李箱、手提袋、宠物等,并增加一个目标轮廓几何约束,来排除轮廓较小或者较大的目标轮廓。上述两个约束的筛选提高了系统对行人目标的辨识率。故本方法在提高标量特征运算鲁棒性的同时,又保证了系统的实时性。

1 算法流程

1.1 图像的获取与技术流程图

笔者利用 2 个 CCD 相机(采取相互平行的拍摄方式)来实现图像和视频的采集。目标识别的技术流程如图 1 所示。

1.2 摄像机的标定

摄像机标定过程涉及 4 个坐标系:世界坐标系、摄像机坐标系、图像物理坐标系和图像像素坐标系^[8],它们之间的位置关系如图 2 所示。

(1)图像像素坐标系 O_0UV 。该坐标系以图像左上角 O_0 为原点,某个像素的坐标 (u,v) 为该像素在图像上所对应的列数与行数,坐标轴的单位为像素。

(2)图像物理坐标系 O_1XY 。将摄像机光轴与摄像机成像平面的交点作为原点 O_1 (图像像素坐标系的中心位置),坐标轴的单位为 mm。图像物理坐标系主要用来描述点在图像中的物理位置。

(3)摄像机坐标系 $O_cX_cY_cZ_c$ 。将摄像机的光心 O_c 作为原点,摄像机的光轴作为 Z_c 轴,平面 $O_cX_cY_c$ 与摄像机成像平面平行,坐标轴的单位为

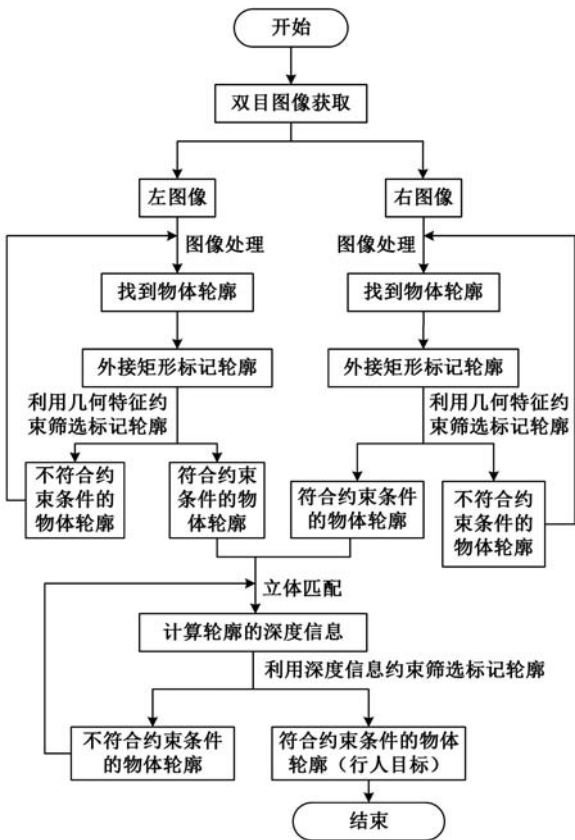


图 1 技术流程图

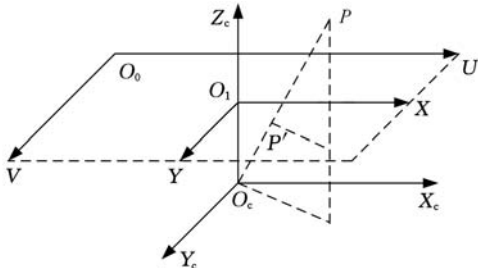


图 2 坐标系之间的关系

mm。摄像机光心到成像平面的距离称为摄像机的焦距 f , $f = O_cO_1$ 。

(4)世界坐标系 $O_wX_wY_wZ_w$ 。世界坐标系用来描述物体的具体位置,可以根据实际要求,设定世界坐标系的原点与方向,坐标轴的单位为 mm。

略去详细的数学推导,图像像素坐标系与世界坐标系之间的转换关系可表示为

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u & 0 \\ 0 & f_y & v & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (1)$$
$$f_x = f/d_x \quad f_y = f/d_y \quad \mathbf{0}^T = [0 \ 0 \ 0]$$

其中,单位像素的尺寸(mm×mm)为 $d_x \times d_y$, O_1 在图像像素坐标系中的坐标为 (u_0, v_0) 。 \mathbf{R} 和 \mathbf{T} 由摄像机坐标系和世界坐标系的相对位置关系决定, \mathbf{R} 为 3×3 的正交旋转矩阵, \mathbf{T} 为 3×1 的平移向量。根据三角测量原理, Z_c 可通过计算两幅

图像的视差得到。 f_x 、 f_y 、 u_0 、 v_0 均只与摄像机的内部结构有关,称为内部参数。 R 和 T 由摄像机坐标系和世界坐标系的相对位置关系决定,称为外部参数。

1.3 几何信息约束的特征提取

本文以行人为研究对象,通过图像处理,提取行人的外部轮廓,将原来复杂的图像简化成便于识别的特征信息,并利用约束条件,对图像中的轮廓进行分类、识别和处理。

1.3.1 运动物体目标提取

本文采用运动物体目标检测算法来识别行走中的行人。最常见的运动目标检测算法有光流法、帧差法、背景减法。背景减法的主要思想是建立一个背景模型,将视频中的每帧图像减去背景模型,得到该帧图像与背景模型之间的差异,即找出运动的目标^[9]。该算法对背景的稳定性要求较高,但算法的复杂性相对较低。程序的工作场景为地铁站,地铁站内的照明条件良好,背景模型的稳定性较高。故本文采取背景减法来检测运动目标,降低算法的复杂度,缩短算法的计算时间。

本文将运动物体目标提取分为以下几个步骤:

(1)根据算法建立背景模型。由于场景并非静止不变,故将背景模型按照一定的速率进行更新,利用高斯滤波处理背景,提高背景模型的鲁棒性^[10]。

(2)读取视频中的一帧图像,通过对图像进行处理,减弱噪声对图像的影响。

(3)将当前帧的图像减去背景模型,得到只含有运动目标的图像,并对其进行二值化处理。

(4)为了减小二值化处理后噪声的干扰,再次利用图像处理来优化图像。

1.3.2 几何信息约束的图像特征提取

本研究中先采取形态学操作来处理图像,再用矩形来标记图像中的每个轮廓,最后利用几何约束来筛选矩形所代表的轮廓。

形态学运算的基本思想是用具有一定形态的结构元素找到图像中的对应形状,以达到图像分割识别的目的,其基本操作有膨胀、腐蚀、开启和关闭。在人群拥挤的情况下,若直接对每个行人进行轮廓提取,则当两个人挨在一起时,系统会误将两人认作一人,从而降低算法的精确度^[6]。本文通过形态学处理,简化每个目标的基本形状,提高算法精度。

用于标记的矩形为图像中轮廓的外接矩形,本文通过外接矩形的参数来描述每个轮廓的相关

信息。轮廓的相关信息包括轮廓的相对位置坐标和轮廓的相对大小。本文设置几何信息约束,利用几何信息约束对目标进行筛选,将处于规定面积范围内的轮廓信息提取出来。目标特征提取的流程分为以下几个步骤:

(1)对图像进行形态学操作,通过形态学操作来填补待测目标上的空洞,让待测目标变得更加完整。

(2)对图像进行形态学腐蚀操作,通过腐蚀操作减小待测目标之间发生干涉的可能性。

(3)用外接矩形依次标记图像中的轮廓曲线,记录外接矩形的相关参数。

(4)利用设置阈值来筛选外接矩形的大小,进而判断该轮廓曲线是否为行人的头部轮廓曲线。

1.4 立体匹配

立体匹配是指在不同摄像机所采集的素材中找到同一目标所对应的特征点。在对图像特征点进行匹配时,需要添加约束准则,一般采用的约束准则有:极线约束、唯一性约束、相似性约束、顺序性约束。

由于实际场景中的待匹配轮廓不多,故本文采用顺序性约束。左侧摄像机拍摄的图像称为左图像,右侧摄像机拍摄的图像称为右图像。顺序性约束采取的准则:用数字依次标记左右图像中的每个外接矩形框,标记的顺序遵循从上到下、从左到右的规律,左右图像中标记数字相同的特征点被认为是来自同一目标。

1.5 深度信息约束

实际应用场景存在许多除行人以外的物体,如手提袋、行李箱、婴儿车等,这些物体会对系统的精确度造成影响。由于这些物体与行人头部存在高度上的差别,故本文利用双目视觉系统来计算图像中每个轮廓的深度信息,通过深度信息对图像中的轮廓进行筛选,提取处于某一高度范围内的目标轮廓。

通过立体匹配,我们可以得到同一目标在左右相机中的位置坐标,将该坐标代入式(1)进行计算,即可获取目标相对于相机安装位置的深度信息。再用相机安装位置的高度减去该目标的深度信息,得到该目标的实际高度。设置深度信息约束,利用深度信息约束对目标高度进行筛选,将处于规定高度范围内的轮廓信息视为行人的头部轮廓。

2 实验数据及分析

本文实验采用微视公司的 MSS2 CCD 摄像

头(2 个),通过相机拍摄图像的方式模拟从视频帧中获取某帧图像,以对算法进行初步的验证,最后用实时采集视频的方式来验证算法的稳定性。

本文利用张友正标定法^[10],选用棋盘平面靶标,该标定板的参数如下:每个格子的规格为 15 mm×15 mm。在 MATLAB 中利用 TOOL-BOX_calib 标定工具箱实现相机的标定。左相机的内参数矩阵为

$$\begin{bmatrix} 876.5 & 0 & 681.3 & 0 \\ 0 & 889.9 & 554.4 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

右相机的内参数矩阵为

$$\begin{bmatrix} 879.1 & 0 & 677.7 & 0 \\ 0 & 890.3 & 522.0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

对比左右相机的内参数矩阵可以看出,左右相机的焦距近似相等。实验图像的分辨率(单位为像素)为 1280×1024,相机的主点坐标(光心坐标)的理论值(单位为像素)应为(640,512),左右相机的主点坐标的误差约为 30 像素。

以下为标定相机所得外参数,其中,旋转矩阵为

$$\begin{bmatrix} 1.0000 & 0.0071 & 0.0043 \\ -0.0073 & 0.9993 & 0.0152 \\ -0.0041 & -0.0374 & 0.9993 \end{bmatrix}$$

平移矩阵为

$$[-82.34 \quad -1.01 \quad -2.90]^T$$

由于采取的拍摄方式为两相机平行拍摄,故外参数的旋转矩阵理论值为 3 阶单位矩阵,从标定结果上看,该旋转矩阵近似等于 3 阶单位矩阵。平移矩阵中第一个参数应为两相机的安装距离,另外两个参数应为 0,即外参数的理论平移矩阵为 $[x \quad 0 \quad 0]^T$ 。从标定结果上看,平移矩阵同样接近理论值,故相机标定的实验结果较为理想。

为了测试算法的性能,在 VS2010 平台下进行了算法的测试。图 3 所示为背景模型,图 4 为实时采集的图像。



图 3 背景模型



图 4 实时采集图像

由图 5 可以看出,通过背景减法可以提取出场景中的目标。由于本文旨在提取行人头部的轮廓,故通过提高相机的曝光量来减小外界光线对

图像产生的影响。从图 6 可以看出,形态学处理后,图 5 中的空洞填满了,并且图像中的轮廓更加平滑,减小了轮廓之间发生干涉的可能性。图 7 所示为利用 canny 算子提取的目标轮廓,通过轮廓的外接矩形来标记每一个轮廓,并用外接矩形的参数来描述目标轮廓的相关信息。外接矩形的参数如表 1 所示。由表 1 可以看出,轮廓 2 的矩形框面积远大于其他轮廓。从图 8 可以看出,通过几何约束的筛选,本方法可以将不符合几何信息约束条件的轮廓 2 排除。

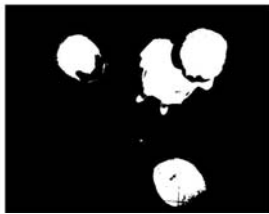


图 5 背景减法处理结果



图 6 形态学处理结果

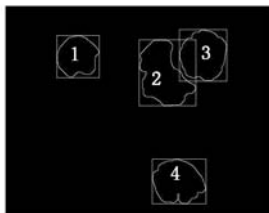


图 7 提取轮廓边界结果

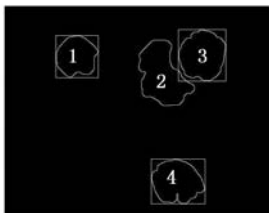


图 8 几何约束筛选结果

表 1 目标轮廓参数表

轮廓	矩形框中心坐标(像素,像素)	矩形框面积(像素 ²)
1	(171,113.5)	9900
2	(382.5,157)	20 748
3	(466.5,116.5)	13 673
4	(409.5,412)	13 462

通过立体匹配,我们可以若判断出同一目标在左右图像中的坐标,将 2 个坐标代入式(1)进行计算,获得目标在世界坐标系下的相对位置坐标,通过世界坐标即可获取该目标的深度信息,如表 2 所示。

表 2 计算世界坐标系下相关参数

轮廓	左相机坐标 (像素,像素)	右相机坐标 (像素,像素)	行人头部的世界坐标 (mm,mm,mm)
1	(171,113.5)	(96.5,127.5)	(176.2,233.4,516.8)
3	(466.5,116.5)	(372,131.5)	(-11.3,241.2,503.6)
4	(409.5,412)	(343.5,432)	(33.7,-69.4,582.4)

笔者将世界坐标系的原点设在左相机的光心处。世界坐标系的坐标轴方向与相机坐标系的坐标轴方向一致,通过计算目标在世界坐标系下的坐标,获取目标相对于左相机安装位置的位置信息。相机的安装高度约是 2300 mm,减去 3 个行人头部世界坐标的 Z 轴坐标 516.8 mm、503.6 mm 和 582.4 mm,计算出参与测试人员的身高为

1783.2 mm、1796.4 mm 和 1717.6 mm。测试人的实际身高约为 1.78 m、1.81 m 和 1.70 m，故本方法可以较准确地获取目标的深度信息。实验欲提取位于高度为 1.75~1.85 m 的轮廓信息，故将深度信息约束的范围设置为 450~550 mm。从图 9 可以看出，通过深度信息的筛选，本方法可以将不符合深度信息约束条件的轮廓 4 排除。

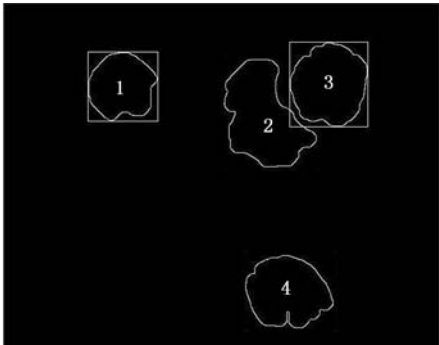


图 9 深度信息筛选结果

为了检测本方法的稳定性，在实验室内对算法进行了测试，实时采集视频，模板检测法与本文检测方法的检测结果如表 3 所示。

表 3 两种方法比较结果

	模板检测法	本文方法
实际行人人数	52	52
算法统计人数	44	47
检测率(%)	84.6	90.4
检测总时间(ms)	162.33	110.26

3 结语

本文利用 2 个相机同时观测物体时存在视差的原理来得到物体的深度信息，通过深度信息和几何特征信息的共同约束将行人头部轮廓从图像中识别出来。通过对比实验可以看出，本方法比模板检测法更省时，满足了实时检测的要求。但该方法对只有部分身体位于相机视场下的行人还存在一定的误报。因此，在后续的研究中，将着重测试该方法在实际场合下的有效性，使其更加贴近实际应用。

参考文献

[1] 关胜晓.机器视觉及其应用发展[J].新型仪器与智能仪表,2005(9):88-92.
Guan Shengxiao.Machine Vision and Its Application [J]. New Type of Instrument and Intelligent Instrument,2005(9):88-92.

[2] 刘润莉.地铁运营客流量计算模型研究[D].成都:电子科技大学,2012.

[3] 顾德军,伍铁军.一种基于人头特征的人数统计方法研究[J].机器制造与自动化,2010,39(4):134-138.

Gu Dejun, Wu Tiejun. Research of Statistical Method Based on the Number of Head Features[J]. Machine Building & Automation, 2010, 39 (4): 134-138.

[4] 顾骋,钱惟贤,陈钱,等.基于双目立体视觉的快速人头检测方法[J].中国激光,2014,42 (1):1-6.
Gu Cheng, Qian Weixian, Chen Qian, et al. Rapid Head Detection Method Based on Binocular Stereo Vision[J]. Chinese Journal of Lasers, 2014,42 (1): 1-6.

[5] 陈双玉,王海辉,吴云韬,等.基于双目立体视觉和 SVM 算法行人检测方法[J].华中科技大学学报, 2015,43(增刊 1):140-143.
Chen Shuangyu, Wang Haihui, Wu Yuntao, et al. Approach of Pedestrian Detection Based on Binocular Stereo Vision and SVM Algorithm[J]. Journal of Huazhong University of Science and Technology, 2015,43(S1):140-143.

[6] Zhao L, Thorpe C E. Stereo and Neural Network-based Pedestrian Detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2000, 1(3): 148-154.

[7] 杨英,刘卫国,钟令,等.基于 AdaBoost 行人检测优化算法的研究 [J]. 机电工程, 2014, 31 (10): 1347-1351.
Yang Ying,Liu Weiguo,Zhong Ling,et al. Based on AdaBoost Pedestrian Detection Research of Optimization Algorithm[J]. Mechanicaland Electrical Engineering,2014,31(10):1347-1351.

[8] 李鹏,王军宁.摄像机标定方法综述[J].山西电子技术,2007(4):78-79.
Li Peng,Wang Junning. Overview of Camera Calibration Methods[J].Shanxi Electronic Technology, 2007(4):78-79.

[9] 王奎奎,玉振明.融合背景减法和帧差法的运动目标检测[J].视频应用与工程,2015,39(24):94-99.
Wang Kuikui, Yu Zhenming. Moving Object Detection Combining Background Subtraction and Frame Difference[J].Video Application & Project, 2015,39(24):94-99.

[10] Zhang Zhenyou, Deriche R,Faugeras O. A Robust Technique for Matching Two Uncalibrated Images through the Recovery of the Unknown Epipolar Geometry [J]. Artificial Intelligence, 1995, 78 (1/2):87-119.

(编辑 张 洋)

作者简介:杨志光,男,1991 年生。武汉轻工大学机械工程学院硕士研究生。主要研究方向为机器视觉。张永林,男,1962 年生。武汉轻工大学机械工程学院教授、博士研究生导师。谢经明,男,1965 年生。华中科技大学机械科学与工程学院副教授。陈幼平,男,1957 年生。华中科技大学机械科学与工程学院教授、博士研究生导师。