

基于双目立体视觉的快速人头检测方法

从大的方面说明为了某种需求提了某个算法

顾 骋 钱惟贤 陈 钱 顾国华 任建乐

(南京理工大学电光学院, 江苏 南京 210094)

两句话概括出本方法整个过程

摘要 为满足人流统计实时性的要求,提出一种人头检测算法。该方法对双目相机采集的图像通过运动目标检测分离出运动人员所在区域,对这些区域使用融合区域匹配和特征匹配的快速匹配方法,即利用视差的连续性只对强纹理点进行绝对误差累积(SAD)匹配,其他点只进行简单的视差验证,能够得到稠密的视差图,再由三角投影关系计算出深度图。由于双目立体成像得到的深度图中人员与场景的深度分布不同,可以采用深度分层的方法将存在人头信息的深度层提取出来,经过数学形态学预处理再利用边缘检测会得到许多候选轮廓,最终利用轮廓的几何特征来判断轮廓是否为人头。实验表明:该算法可以很好地适应复杂场景下的人头检测,精度高、速度快。

关键词 图像处理; 人头检测; 双目立体视觉; 绝对误差累积匹配; 视差图

中图分类号 TN29 **文献标识码** A **doi**: 10.3788/CJL201441.0108001

用实例体现本方法的效果

Rapid Head Detection Method Based on Binocular Stereo Vision

Gu Cheng Qian Weixian Chen Qian Gu Guohua Ren Jianle

(College of Optoelectronic Technology, Nanjing University of Science and Technology,
Nanjing, Jiangsu 210094, China)

Abstract To meet the requirements of real-time statistics of the stream of people, a head detection algorithm is presented. The method isolates the region with moving persons by the moving target detection in the binocular camera images, a rapid matching method mixed region matching and feature matching is used to this region. Using the continuity of disparity, we just need to match the strong texture points by sum of absolute differences (SAD) method and simply verify the disparity to other points, so we can get the dense disparity map, and then calculate the depth map by the triangle projection relation. As the depth distributions between people and scene are different in the depth map generated by binocular stereo imaging, we can use the mathematical morphology operations to precondition, then get many candidate contours by edge detection. The characteristics of the head are used to determine whether the contour is the head. Experiments show that this algorithm can be well adapted to head detection in complex scenarios with high precision and speed.

Key words image processing; head detection; binocular stereo vision; sum of absolute differences matching; disparity map

OCIS codes 100.2000; 100.6890; 140.4065; 140.6910

1 引言

在许多商业和公共场所,如车站、公共汽车、商场、影院等都需要采集人员流动信息。这些信息能够用来决定商场的租赁价格、商品的销售情况、广告投放的效益;帮助车站安全高效地安排附属设施;帮助交通运营部门恰当地分配人力、物力、财力,合理

地调整路线。出于上述原因,自动人员计数系统一直是研究人员关注的热点。随着计算机视觉的兴起,其可视化、自动化、对复杂场景的高适应性等优点将自动人员检测系统引入了新的领域,发展出了许多高效的检测方法。这些方法有一个共同点就是通过检测人头来实现流动人员的监测。

收稿日期: 2013-08-12; 收到修改稿日期: 2013-09-03

基金项目: 国家自然科学基金(61271332)、江苏省“六大人才高峰”支持计划(2010-DZXX-022)

作者简介: 顾 骋(1988—),男,硕士研究生,主要从事图像处理和机器视觉等方面的研究。E-mail: 462307589@qq.com

导师简介: 钱惟贤(1980—),男,副研究员,硕士生导师,主要从事目标探测方面的研究。

E-mail: developer_plus@vip.163.com

常用的人头检测方法有三种:一是基于轮廓的检测技术,头部轮廓特征主要由头部边缘点的梯度参数构成,常用的轮廓模型有椭圆模型、高斯模型、曲线模型^[1];二是基于颜色的检测技术,主要有发色模型和肤色模型^[2];三是基于模板的检测技术,利用特征模板或面部模板检测人头,需要模板库、训练器。基于轮廓的检测方法对图像分辨率低、人物角度多样、遮挡较多的情况稳健性较差;基于颜色的检测方法容易受光照影响,且分离颜色时需要经常调整阈值,实用性不佳;基于模板的检测方法具有较好的稳健性,正确率也较高,但是分类器的训练比较耗时,当特征量比较多时会增加空间复杂度,难以满足实时监控的需求^[3]。

上述的方法都是基于单目摄像机的,主要的不足之处在于遮挡、外形、光照、分类器训练复杂。为了解决这些问题,本文提出了一种基于双目立体视觉的快速人头检测方法,对拍摄的左右视图通过立体匹配以获取场景中物体相对于摄像机的深度图,对深度分层提取出含有人头的那一层深度,再通过进一步的数学形态学操作和轮廓检测就能够提取人头。该方法利用双目立体视觉高精度、对光照的不敏感的特点,实现了复杂场景的人头检测。

2 双目立体视觉

2.1 立体成像原理

立体视觉兴起于 20 世纪 60 年代,近年来一直是计算机视觉领域最热门的研究问题之一。立体视觉的基本原理就是通过模拟人类视觉处理景物的方式,在不同的位置上同时拍摄两幅图像,通过三角测量原理计算两幅图像上像素点间的位置偏差,获取物体的深度信息。经过多年的研究,双目视觉理论不断发展,并且由于软硬件性能的不断提高,立体视觉发挥着越来越大的作用,目前已广泛应用于工业检测、机器人导航中^[4]。

图 1 显示的是一套无畸变、行对准的标准立体试验台^[5](通过图像校正可以调节两摄像机之间的位置关系,使其接近于标准立体模型,使得左右焦距 f_l 、 f_r 和主点坐标 C_l 、 C_r 都完全相同)。其中 C_l 、 C_r 分别是左右视图的主点(区别于图像的中心点,是光轴与像平面之间的交点), O_l 、 O_r 分别是左右摄像机镜头透镜的光心, T 是基线距离(两摄像机的间距),物理场景中的一个点 P 投影在左右视图上,由于校正后的图像满足极线定理,所以点 P 在左右视图上成像点的纵坐标值是完全相同的,不同的则是横坐标

值,分别为 x_l 、 x_r ,记 $d = x_l - x_r$,其中 d 是视差,再根据相似三角形原理就能推导出点 P 距离双目摄像机的深度 Z ^[6]。

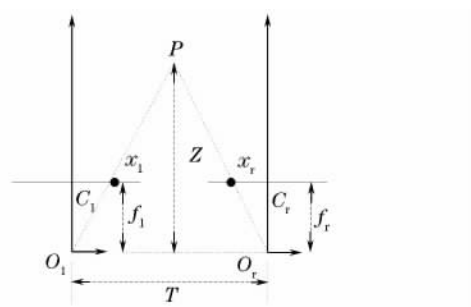


图 1 无畸变、行对准的标准立体试验台

Fig. 1 Undistorted and line-aligned standard stereo test bed

由相似三角形原理,可得到

$$\frac{T - (x_l - x_r)}{Z - f} = \frac{T}{Z}, \quad (1)$$

$$Z = \frac{fT}{x_l - x_r}, \quad (2)$$

式中 f 是双目相机的焦距($f = f_l = f_r$),结合视差表达式继而可以得到深度精度

$$\Delta Z = \frac{Z^2}{fT} \Delta d, \quad (3)$$

式中 Δd 是图像的精度,单位是像素,对于给定的摄像机参数,提高场景点深度计算精度的有效途径是增长基线距离和降低 Δd 。

2.2 立体匹配

立体视觉中最为重要的就是立体匹配,立体匹配效果的好坏直接影响到图像处理最终结果的好坏^[7]。立体匹配一直是一个“病态”问题,到现在也没有一种方法可以完美地解决图像对应点的问题,大多只能单方面地顾及实时性和准确性^[8]。立体匹配根据约束信息的不同可以分为局部算法和全局算法。局部算法利用的是对应点本身以及附近区域的约束信息,由于其选取的是局部最优解,所以对一些由于遮挡和单一纹理造成的模糊比较敏感,但优点是效率高,根据匹配基元的不同可以分为区域匹配、特征匹配、相位匹配;全局算法利用了图像的全局约束信息,对局部的模糊不敏感,缺点是计算代价很高,全局算法一般有动态规划的算法和图切割的算法^[9]。

对于人头检测而言,需要的是在一定准确性的前提下努力地提高实时性,所以立体匹配选取局部算法是很好的选择。文献^[10]采用的是一种称为绝

对误差累积(SAD)的区域匹配方法来进行立体匹配,获得了稠密的视差图,再通过分割的方法提取出人头,实现了不错的检测效果。但其处理时需要每一个像素点都进行立体匹配,这样就大大降低了实时性。为了兼顾实时性和准确性,这里提出了一种融合区域匹配和特征匹配的快速立体匹配方法。

该方法是基于视差的分段连续性,由于场景中大部分区域的视差都是连续的。所以这里首先用Canny边缘检测提取出强纹理点,对强纹理点进行SAD匹配求视差;对非强纹理点则进行简单的视差验证,满足验证条件的则视差值不改变,不满足条件的则再次使用SAD匹配求视差。验证的方法就是将上一个点的视差值代入,计算出左右图像在此视差值下的SAD值,若小于阈值 T_{th} 说明视差未发生

跳变;若大于阈值 T_{th} 说明视差存在跳变,需当重新计算视差。流程如图2所示。

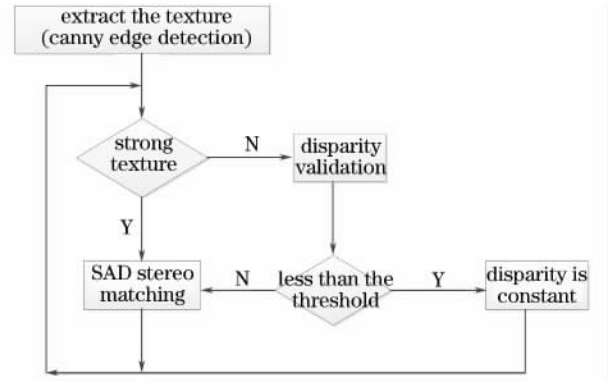


图2 快速立体匹配流程图

Fig. 2 Flow chart of rapid stereo matching

SAD相似性测度函数表示为

$$C_{sad}(x, y, d) = \sum_{i=-N}^N \sum_{j=-N}^N |I_r(x+i, y+j) - I_l(x+d+i, y+j)|, \quad (4)$$

$$d_{sad}(x, y) = \min \{C_{sad}(x, y, d), d \in [d_{min}, d_{max}]\}. \quad (5)$$

一般情况下,两个摄像机的特性是不同的,同一场景处的光强可能存在一定的差异。若是直接进行SAD相似性匹配,得到的结果会受到摄像机特性的影响,因此需要进一步归一化处理,得到

$$C_{nsad}(x, y, d) = \frac{\sum_{i=-N}^N \sum_{j=-N}^N |I_r(x+i, y+j) - I_l(x+d+i, y+j)|}{\sqrt{\sum_{i=-N}^N \sum_{j=-N}^N I_r(x+i, y+j) \sum_{i=-N}^N \sum_{j=-N}^N I_l(x+d+i, y+j)}}, \quad (6)$$

$$d_{nsad}(x, y) = \min \{C_{nsad}(x, y, d), d \in [d_{min}, d_{max}]\}. \quad (7)$$

3 算法的原理和步骤

3.1 人头检测算法的总体框架

整体流程如图3所示,主要包括运动目标检测、深度图分层、人头轮廓提取三大部分。

3.2 运动目标检测

对于整个场景而言,这里所感兴趣的只是运动的人员,背景信息可看作冗余信息,而且背景的存在对于人头的检测还会存在一定程度的干扰,因此就运用到了运动目标提取的技术。提取运动目标有两点好处:一是防止背景对人头分割的影响;二是由于只需要对运动物体区域进行立体匹配,计算所消耗的时间也大为减少。

这里使用融合背景差分法和帧间差分的方法对运动目标进行检测,使两种方法优势互补^[11]。首先连续采集 N 帧图像取平均来得到原始背景 B_0 :

$$B_0 = \frac{1}{N} \sum_{k=0}^{N-1} I_k(x, y), \quad (8)$$

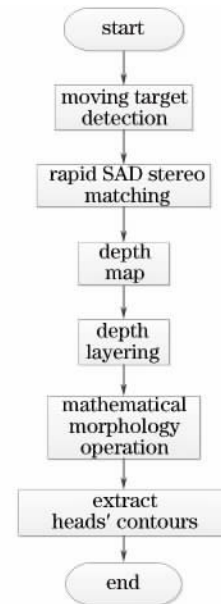


图3 算法总体流程图

Fig. 3 Flow chart of the overall algorithm

式中 $I_k(x, y)$ 是采集到的第 k 幅图像, 使用一阶递归滤波器对背景图像动态更新, 得到

$$B_k(x, y) = \alpha I_k(x, y) + (1 - \alpha) B_{k-1}(x, y), \quad (9)$$

式中 $B_k(x, y)$ 表示当前背景图像, $B_{k-1}(x, y)$ 表示上一帧背景图像。递归系数 α 表示当前帧图像在背景图像中的贡献度, α 越大说明当前帧对背景图像的贡献度越大, 该运动目标检测法的性能越接近于帧间差分法, 适合于探测运动速度较快的物体; α 越小说明当前帧对背景图像的贡献度越小, 该运动目标检测法的性能更接近于背景差分法, 适合于探测运动速度较慢的物体, 文中选取的 α 值为 0.03。

对背景图像和当前帧图像做帧间差分, 以阈值 T 对帧差图像二值化处理, 就可以得到运动目标灰度值^[12]

$$D(x, y) = \begin{cases} 0, & |I_k(x, y) - B_{k-1}(x, y)| < T \\ 255, & |I_k(x, y) - B_{k-1}(x, y)| > T \end{cases} \quad (10)$$

3.3 深度图分层

对运动区域使用提出的快速立体匹配方法可以得到视差图。在焦距、基线距离一定的情况下, 视差的搜索范围则完全由摄像机到探测目标的距离决定, 距离越近则搜索的范围越宽, 越远则搜索的范围越窄, 若能预先知道探测的范围就能大大缩小搜索范围, 减少不必要的时间消耗。在摄像机与地面距离已知的情况下, 通过限定人员的身高范围就可以确定视差搜索的左右极限

$$d_{\max} = fT/Z_{\text{low}}, \quad (11)$$

$$d_{\min} = fT/Z_{\text{high}}, \quad (12)$$

式中 d_{\max} 表示搜索的右极限(最大视差), d_{\min} 表示搜索的左极限(最小视差), Z_{low} 表示摄像机到地面的距离减去人员的最高高度, Z_{high} 表示摄像机到地面的距离减去人员的最低高度。实验中摄像机到地面的高度是 2.7 m, 人员的高度限定在 1.3 ~ 2.0 m, 所以摄像机距离人员的高度在 0.7 ~ 1.4 m, 通过(11)、(12)式可以推算出视差搜索范围为 20 ~ 40 pixel(摄像机基线距离为 60 mm, 相机焦距为 466.67 pixel)。

但仅仅一幅视差图并不能满足检测人头的要求, 还需要能够反映物体与摄像机之间距离信息的深度图。通过三角变换就可以将视差图转换为深度图, 当然为了计算的方便还需要将其归一化到 0 ~ 255 的范围内。

深度图呈现出的是运动人员的整个深度信息, 其中包含了最终需要检测的人头信息。利用人体的

身高特征, 头部肯定是整个身体中距离地面最远的, 也即距离摄像机传感器最近的部分。所以通过对深度图分层就可以把含有人头信息的那一段深度截取出来单独处理。

3.4 人头轮廓提取

深度层二值化处理后的图像存在着若干连通区域, 这些区域中存在着孔洞和边缘毛刺, 所以首选的就是数学形态学操作。数学形态学操作可以用来简化图形数据, 提取图像中的基本形态元素, 去掉与研究无关的冗余成分, 在区域分割中有很重要的应用。对二值图像使用膨胀、腐蚀算子进行形态学操作, 然后对其结果图像使用 Canny 边缘检测算子提取出比较完整的边缘轮廓, 最后就可以得到一组候选轮廓序列。

接下来筛选上一步得到的轮廓序列。判断的条件是一些先验知识, 如离散度、占空比、周长、面积。离散度、占空比分别表示为

$$F = L^2/A_i, \quad (13)$$

$$M = A_i/A_o, \quad (14)$$

式中 F 是轮廓的离散度, L 是轮廓的周长, M 是轮廓的占空比, A_i 是轮廓包围区域的面积, A_o 是轮廓外接矩形的面积。实验中取 L 阈值为 40, A_i 阈值为 100, F 阈值为 16, M 阈值为 0.7, 符合这 4 个指标的轮廓就被判定为人头。

4 实验结果与分析

实验中将双目摄像机吊顶安装在实验室房顶上, 使其自上而下拍摄图像。图 4 中(a)~(j)为对多个人头检测的结果, 其中(a)、(b)为摄像机拍摄的左右视图, (c)为 Canny 边缘检测提取出的强纹理点, (d)为使用了融合背景差分法和帧间差分的方法提取出的运动区域, (e)为使用了所提快速立体匹配方法后得到的原始深度图, 在深度图中可以看到明显的分层效果, 离摄像机最近的头部像素值最高, 而离摄像机最远的地面像素值最低。通过对深度图分层, 就可以得到包含人头信息的那一层深度如(f)所示, 如何选取截取的深度需要根据实际场景多次实验获得, (f)中将 1.3 ~ 1.6 m 的深度层截取掉, 保留的是 1.6 ~ 2.0 m 的深度层, 此时白色区域附近含有很多的孔洞和毛刺, 需要使用形态学膨胀腐蚀操作使图像变得更加平滑, (g)中人头轮廓已经比较明显, 使用 Canny 边缘检测即可将轮廓提取出来, (h)中获得的轮廓组成了一组轮廓序列, 对所有轮廓序列通过先验知识进行筛选, 最终只有头部的轮廓是

符合条件的,(i)为筛选出的头部轮廓并对其进行了填充,(j)根据轮廓的位置在原图中用红色圆圈标记了出来。

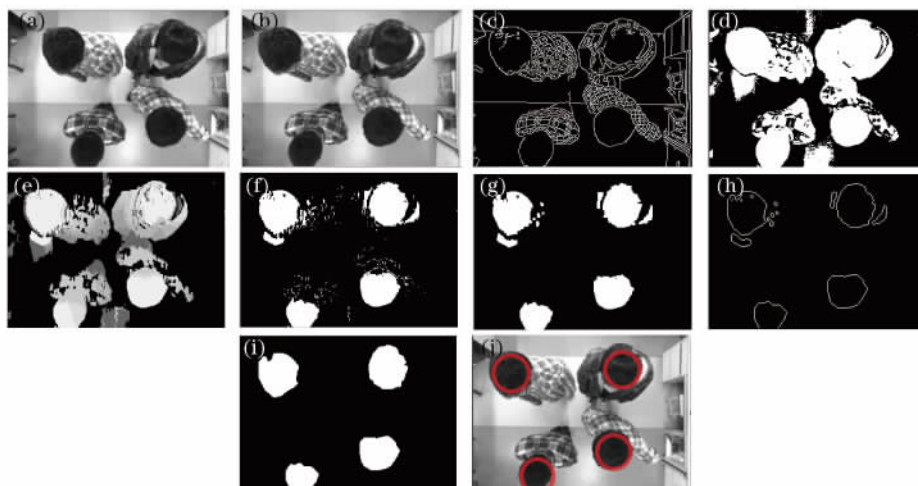


图4 (a)左视图;(b)右视图;(c)边缘检测;(d)运动目标提取;(e)原始深度图;
(f)分层后的深度图;(g)形态学操作;(h)边缘检测;(i)轮廓提取并填充;(j)人头检测结果

Fig.4 (a) Left view; (b) right view; (c) edge detection; (d) extract moving objects; (e) original depth map; (f) depth map after stratification; (g) morphological operations; (h) edge detection; (i) contour extraction and filling; (j) result of heads' detection

为了进一步提高算法的速度,可以开辟多线程来并行处理数据,表1显示的是分别用文献[10]方法、所提方法、所提方法(多线程)的运行时间,从表中显示的结果可以清晰地看出通过使用所提快速立体匹配的方法使得整个人头检测程序运算的时间提

高一个量级,而开辟多线程之后速度又提高了不少,使得程序可以实时处理数据。(表1数据是在i5处理器、4 GB内存、32bit Windows7操作系统的笔记本上测算出来的)

表1 算法实时性比较

Table 1 Comparition of algorithm's real-time property

Stereo matching method	Time for stereo matching /s	Time for heads detection /s
Ref [10]'s method	7.107	7.157
Proposed method	0.180	0.231
Proposed method (multithread)	0.101	0.130

5 结 论

基于双目立体视觉的方法,利用双目同时观测物体时会存在视差的原理来得到整个场景的深度图,将含有人头信息的深度层分离出来再利用轮廓检测识别出人头。该方法精度高、抗干扰性好,对于复杂场景中的人头可以有效地检测,通过所提的方法可以实时地处理数据,具有很强的实用性和推广性。在进一步的研究中将着重测试人头检测系统在不同场景中的适应性,以增强算法的实用性。

参 考 文 献

- 1 Zhang Haiyang, Chen Guolong, Li Xianwei. Research of head detection based on curvature scale space algorithm[J]. Computer Engineering and Applications, 2012, 48(14): 195-197.
- 2 Zhao Junwei, Hou Qingtao, Li Jinping, *et al.*. Head detection algorithm based on mathematical morphology and HSI color space [J]. Journal of Shandong University (Engineering Science), 2013, 43(2): 5-10.
- 3 He Yangming, Dai Shuguang. Application of clustering and technology[J]. Computer Engineering and Application, 2009, 45(31): 145-146.
- 4 Zhang Xuping, Wang Jiaqi, Zhang Yixin, *et al.*. Large-scale three-dimensional stereo vision geometric measurement system [J]. Acta Optica Sinica, 2012, 32(3): 0315002.
- 5 Gray Bradski, Adrian Kaebler. Learning OpenCV [M]. Yu Shiqi, Liu Ruizhen, Transl. Beijing: Tsinghua University Press, 2012, 48(14): 195-197.

- 2009, 452—455.
- 格 瑞, 安德里安. 学习 OpenCV[M]. 于仕琪, 刘瑞祯, 译. 北京: 清华大学出版社, 2009. 452—455.
- 6 He Bingwei, Lin Dongyi, Chen Zhipeng, *et al.*. Research of eliminating occlusion in visual construction of three-dimensional objects[J]. Chinese J Lasers, 2011, 38(7): 0708001.
- 何炳薇, 林东艺, 陈志鹏, 等. 三维物体视觉测量重构中解决遮挡问题的方法研究[J]. 中国激光, 2011, 38(7): 0708001.
- 7 Hou Junjie, Wei Xinguo, Sun Junhua. Calibration method for binocular vision based on matching synthetic images of concentric circles[J]. Acta Optica Sinica, 2012, 32(3): 0315003.
- 侯俊捷, 魏新国, 孙军华. 基于同心圆合成图像匹配的双目视觉标定[J]. 光学学报, 2012, 32(3): 0315003.
- 8 Yu Naigong, Qin Yonggang, Ruan Xiaogang. Review of stereo correspondence algorithms [J]. Computer Measurement & Control, 2009, 17(5): 817—819.
- 于乃功, 秦永刚, 阮晓刚. 立体匹配算法进展[J]. 计算机测量与控制, 2009, 17(5): 817—819.
- 9 Zeng Jiguo, Yu Guohe, Xie Xusheng, *et al.*. Fast and robust stereo matching method [J]. Computer Engineering and Applications, 2013, 49(14): 196—199.
- 曾纪国, 余国和, 谢旭升, 等. 快速鲁棒的立体匹配方法[J]. 计算机工程与应用, 2013, 49(14): 196—199.
- 10 Tarek Yahiaoui, Cyril Meurie, Louahdi Khoudour, *et al.*. A People Counting System Based on Dense and Close Stereovision [M]// Elmoataz A, Lezoray O, Nouboud F, *et al.* Image and Signal Processing. Berlin: Spriger Berlin Heidelberg, 2008, 5099: 59—66.
- 11 Zhen Jiye, Yan Wantao. A moving target detection algorithm based on dynamic background updating[J]. Computer Knowledge and Technology, 2010, 6(13): 3499—3500.
- 郑纪业, 闫万涛. 基于动态背景更新的运动目标检测算法[J]. 电脑知识与技术, 2010, 6(13): 3499—3500.
- 12 Yuan Weicai, Xu Xiangming. Effective algorithm of background extraction and updating [J]. Computer Engineering and Application, 2010, 46(4): 191—192.
- 袁伟才, 徐向民. 一种有效的动态背景提取及更新方法[J]. 计算机工程与应用, 2010, 46(4): 191—192.

栏目编辑:何卓铭