

文章编号 1004-924X(2013)04-1079-07

基于 Mean Shift 图像分割和支持向量机 判决的候梯人数视觉检测系统

张宇洋, 刘满华, 韩 韬*

(上海交通大学 电子信息与电气工程学院, 上海 200240)

摘要:根据电梯群控系统的需求,提出了一种基于视觉检测技术获得候梯人数的新方法。考虑候梯人数检测系统的监测目标为候梯人群,而候梯人群的心理、建筑风格,摄像机的安装角度、复杂背景等因素均会影响到待识别模式的提取,故作者提出了以人体头部作为模式进行模式识别来检测候梯乘客的数量。该方法以 Mean Shift 图像分割算法和支持向量机(SVM)决策分类器为核心,考虑候梯人群图像采集角度、拍摄镜头的特殊性等对候梯人群头部进行精确识别,较为快速地得到了准确的识别结果。实验证明,该方法处理图像速度可保持在每幅图片 2 s 以内,准确率超过 80%,满足了电梯群控系统的需求。由于能够使电梯群控系统获得稳定可靠的输入参数,从而提高了电梯群的运送效率。

关键词:视觉检测; Mean Shift 图像分割; 支持向量机决策分类; 人数检测

中图分类号: TP391.4 **文献标识码:** A **doi:**10.3788/OPE.20132104.1079

Elevator-waiting people counting system based on Mean Shift segmentation and SVM classification

ZHANG Yu-yang, LIU Man-hua, HAN Tao*

(*Institute of Electron Information and Electrical Engineering,
Shanghai Jiao Tong University, Shanghai 200240, China*)

* *Corresponding author, E-mail: than@sjtu.edu.cn*

Abstract: According to the requirements of elevator-group-control systems, a new method based on computer vision detection technology was proposed to obtain the number of passengers waiting outside of an elevator. As the detecting target of the system was the passengers waiting outside of the elevator, the feelings of passengers, building styles, the installation angle of a camera and the complex background would effect the detection pattern. Therefore, this paper took the human head as the model to implement the pattern recognition and to detect the number of passengers waiting outside of the elevator. The proposed method based on computer vision detection algorithm combined Mean Shift image segmentation and Support Vector Machine (SVM) classification and recognized the human head features according to the angles of image acquisition and special lenses of cameras. It can obtain accurate recognition results. Experimental results show that the method has an image processing speed by 2 s/image in real time and the accuracy above 80%, which meets the

收稿日期:2013-01-09;修订日期:2013-01-28.

基金项目:国家教育部博士点基金资助项目(No. 20090073120019)

needs of elevator-group-control systems. As a results, the transport efficiency has been improved greatly due to the stable input parameters for the elevator-group-control systems.

Key words: computer vision detection; Mean Shift image segmentation; Support Vector Machine (SVM) classification; people counting

1 引 言

在多电梯群控系统中,若能提前获知各层候梯人数,将若干电梯优先服务于等候人数较多的楼层,便可以提升电梯群的运送效率,缩短候梯时间。相比于轿厢称重等后验检测方法,基于计算机视觉的人数检测方案更为准确和高效。但候梯人群心理和制造成本等因素限制了摄像机选取以及采集到的图像质量;建筑风格、候梯大厅的场景、光照变化的多样性,使图像背景复杂多变;电梯控制器计算能力和储存器等硬件成本以及检测实时性也限制了检测算法设计时的复杂性。

选择什么目标作为待识别模式是该检测方案中最重要的问题。传统的识别人脸然后计数的方法,如基于启发式模型的镶嵌图(Mosaic Image)人脸检测^[1]及改进的镶嵌法^[2],由于候梯乘客不会全部主动地正面对摄像头,所以并不适用。以人体肩膀间隔作为待识别的模式完成人群分割的方法,只适用于大门入口等场景人数检测^[3],并要求目标人群排列成为平行的一列,对随意站立的候梯人群同样不适用。该项应用考虑到单目俯视图像、候梯大厅的面积和顶高、光照和复杂背景以及人流密集的情况下出现互相遮挡、粘连等问题,选择了头部轮廓以及肤色和发色颜色分布的统计信息等特征作为识别人头区域的依据,不过由于乘客服装、帽子的颜色可能存在与人体皮肤颜色发色相近的情况,简单的彩色直方图模板匹配方法也会引入不可控误差^[4]。文献^[5]提出了在基于梯度信息的 Hough 变换基础上,利用改进的椭圆间距测度和共圆置信度识别出图像中所有真实人体头部并提取出每个人体头部区域的最优拟合轮廓,从而获得人体头部的精确定位方法。该方法利用灰度判决器和基于视知觉分组理论的共圆判决器去除大量的虚假候选头部区域轮廓,具有一定的准确率,但在摄像机拍摄角度不能完全俯视时,头部的椭圆形轮廓特征将没有那么明显,存在漏检的可能;同时,当背景或目标服装中

有物体轮廓与头部轮廓相似时,会产生误检测。

在对可能成为头部目标的区域进行判决之前,必须通过一定的方法将可能为头部的目标区域与背景分离开来。运动检测中的帧间差分法对光线变化不敏感,且检测速度快,常用于目标的运动检测;但该方法一般不能完全提取出所有相关的特征像素点,对于复杂背景下的运动目标检测效果堪忧。背景减法对于复杂背景下的运动目标的检测效果通常优于前者^[3,6],但对于光照和外来无关事件等干扰特别敏感,需要进行适当的改进。如基于卡尔曼滤波的自适应背景模型和自适应的混合高斯背景模型能在一定程度上减少光照等对效果产生的影响。利用图像分割的方法,如基于 graph cut 分割的轮廓检测方法^[7],可对各目标的边界进行提取并准确地与背景分割开来,同时再附加一个判决过程,如包涵图像的边缘的部分为背景,便可以将背景去除;但由于该方法往往只能提取目标的整体轮廓,并不能将头部与目标与身体其他部位分离出来,在多个目标有重叠情况发生时,往往会使后续检测过程有一定误差。

本文提出了基于 Mean Shift 的图像分割算法作为去除背景,提取目标的方法,将分割后面积较大的区域作为背景区域去除,所有面积较小的区域均作为目标区域送入支持向量机(Support Vector Machine, SVM)决策分类器中进行判决。Mean Shift 分割算法既不会像分水岭算法那样形成过度的分割,也不会像基于 graph cut 分割的轮廓提取算法那样提取整个目标。该算法可将多个目标准确地分割开来,全部作为单个目标处理,避免了多目标的判决统计过程,达到了较好的候梯人数统计效果。

2 图像采集

根据上文所述,系统的框图如图 1 所示。

首先我们分析图像采集模块。对于候梯人数检测系统使用的摄像机有如下需求:覆盖范围需要能够包括整个电梯的候梯区域,且尽量避免遮

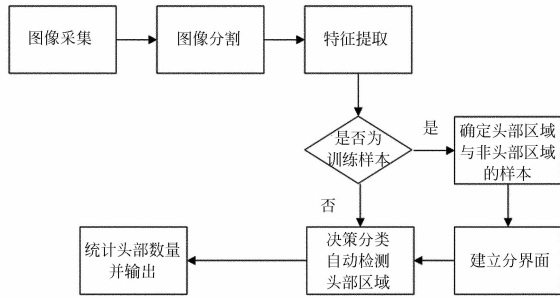


图 1 系统框图

Fig. 1 System diagram

挡等问题;尽可能考虑候梯乘客的感受,不使用较明显的大型枪体摄像机,且摄像机在安装后有一定的隐蔽性;在能够保证检测精度的前提下,尽量控制摄像头的成本。根据电梯生产厂商所提供的数据,候梯区域长宽应保证不小于 4 m×3.5 m,安装高度约为 3.5~5.5 m。

为保证系统工作尽可能稳定可靠,设置视野范围 ROI(Region of Interest)为 4.5 m×4.5 m。为了使 ROI 能够覆盖电梯门口,实际安装摄像机时倾角 θ 约等于 0,图 2 所示。针对这些需求,本系统选择采用覆盖范围较大的半球型广角摄像机,安装于电梯门所在墙一侧,采用自顶向下的拍摄角度。为折衷图像采集质量和摄像头成本,选择 1/3 in(1 in=2.54 cm)的 SONY CCD,有效像素为 752(水平)×582(垂直)。由于采用 1/3 in CCD(感光元件对角线长度 6 mm,582 个像素点在 CCD 中所占物理长度约为 3.67 mm)。为了能保持摄像头的有效视野范围能在 3.5 m 至 5.5 m 的可能安装高度中都能覆盖到完整的候梯人群,应选择焦距可调的摄像机,且可调范围应涵盖 2.85~4.49 mm。

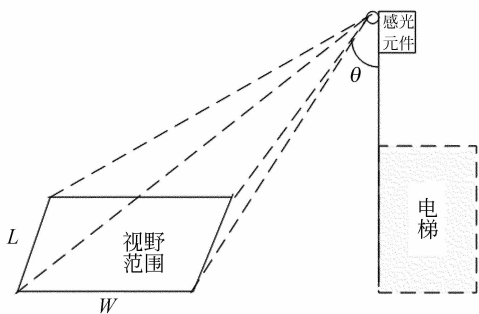


图 2 摄像机视野范围

Fig. 2 ROI(Region of Interest)

满足了上述需求的摄像机,采集过程也不可避免地产生了一些影响检测系统准确性的问题,如部分图像的变形、人脸信息不完整、人脸姿态、方向的不确定等。这些都是本应用需要克服的难点,成为了图像分割与决策分类时不得不考虑的因素之一。

3 图像分割

由于本文提出的人数统计的核心方法为图像分割配合决策分类从而实现对候梯人群的头部进行识别,然而作为决策分类器的输入端,图像分割的准确率将直接影响到决策分类器的准确率。为保证实时性,图像分割算法的执行速度也非常重要。

当 Mean Shift 算法运用于图像分割时,统一考虑图像的空间信息和色彩(或灰度等)信息,组成一个 $p+2$ 维的向量 $x=(x^s, x^r)$,其中 x^s 表示网格点的坐标, x^r 表示该网格点上 p 维向量特征,当图像为灰度图像时, $p=1$;彩色 RGB 图像时, $p=3$ 。分别用 x_i 和 $z_i, i=1, \dots, n$ 表示原始和分割后的图像。定义向量

$$m_h(x) = \frac{\sum_{i=1}^n K(\frac{x_i - x}{h}) \omega(x_i) x_i}{\sum_{i=1}^n K(\frac{x_i - x}{h}) \omega(x_i)}, \quad (1)$$

其中: $\omega(x_i)$ 为采样点 x 的权重,离 x 近的采样点 x_i 有较大的权重,即离 x 越近的采样点对估计 x 周围的统计特性越有效,反之亦然。

定义核函数来 K_{h_s, h_r} 估计 x 的分布,具有如下形式:

$$K_{h_s, h_r} = \frac{C}{h_s^2 h_r^p} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right), \quad (2)$$

其中: C 是一个归一化常数, h_s, h_r 为带宽,带宽的大小可以看作是分割分辨率,带宽越大,越多的图像细节就会被忽略。根据参考文献[8-10],空间带宽 h_s 是分割后两类在空间上的距离,在实际分割过程中表现为空间窗的大小; h_r 为像素点的 RGB 属性带宽(在 RGB 图像中),表示分割后两类在 RGB 属性上的距离。在本应用中,由于我们需要提取到候梯乘客的人头部分区域,需要一定的分割分辨率,所以 h_s, h_r 不宜太小。同时,由于后续的判决器需要人头区域内的肤色以及发色的颜色统计信息,所以要求该分割过程不能将完整

的头部区域分割开,所以, h_s, h_r 不宜选取特别大的值。

用 Mean Shift 算法进行图像分割的具体步骤如下(对每一个像素点):

1. 初始化 $j=1$, 并且使 $y_{i,1} = x_i$

2. 计算 $m_h(x)$

3. 把 $m_h(x)$ 赋给 x

4. 如果 $\|m_h(x) - x\| < \epsilon$, Mean Shift 过程结束, 记收敛后的值为 $y_{i,c}$; 若不然, 继续执行(1)。

5. 赋值 $z_i = (x_i^s, y_{i,c}^r)$

6. 收敛至同一点的起始点归为一类, 合并像素点过少的类, 融合局部像素, 得到分割结果。

最终, 图像中的每一个像素点都有自己所属类的类标, 实现了图像分割。

4 特征提取与决策分类

由于摄像机拍摄角度为自上而下, 检测候梯人数最实际可行的方法便是检测头部的数量。经过图像分割之后, 得到一系列区域, 利用基于机器学习的模式识别方法, 可以建立一个 SVM 决策分类器^[11]。将这些区域分成两类, 一类为是头部, 一类为不是头部。

4.1 特征提取

一块分割后的区域可以在多个方面, 如颜色、边缘、纹理、形状和大小等方面被描述。其中, 彩色直方图作为一种代表目标的色彩统计特征的非常有效的图像描述, 在模式识别以及目标跟踪中应用非常广泛^[12], 而在此处, 彩色直方图信息表示的即为肤色及发色在某个区域内的颜色统计信息, 对于满足人头部肤色发色区域彩色直方图一定阈值范围内的区域, 即可判为人头区域, 因此该特征为关键的判决依据, 但在发色不为黑色或头戴白色帽子等较为极端的情况下, 该特征有失效的可能; 在本应用中, 区域的大小(所占像素点的个数)和形状(长轴长与短轴长的比值)也是非常重要的特征, 考虑与彩色直方图配合使用。选取不同特征在实际测试实验的平均准确率与均方差如表 1 所示, 其中单次准确率=(测试人数个数-实际人数个数)/实际人数个数。其中使用的训练样本为多个候梯厅、一段时间内有电梯乘客使用电梯的约 1 000 幅图片。

表 1 选取不同特征时的准确率

Tab. 1 Accuracy of different features

选用特征	平均准确率	均方差
RGB 直方图+大小+形状	82.49%	0.024 522
RGB 直方图	83.77%	0.020 737
RGB 直方图+大小	83.36%	0.016 951
RGB 直方图+形状	79.36%	0.026 173



图 3 特征选取对比示例图

Fig. 3 Comparison of feature selections

根据表 1 的实验数据, 虽然彩色直方图特征在实际检测中起着决定性的作用, 但加入区域大小特征后, 平均准确率与均方差指标均有所提高, 整体误判率减小, 稳定性也略有提高。究其原因, 由于候梯乘客站姿、面对方向、身高等都不尽相同, 这样的拍摄角度下在图片中头部呈现出的形状(长轴长与短轴长之比)没有很好的聚类特性, 加入形状特征不能提高整体检测效果, 甚至有可能适得其反, 在一定程度上影响分类器的分类效果, 如图 3(a)所示, 由于 1 号目标的头部位置处于广角镜头畸变最严重的位置, 收到非常严重的变形, 若加入形状信息作为特征, 出现了漏检; 而头部大小表现在图像中即为头部所占的像素点个数, 在这样的拍摄角度下不会有太大的浮动, 且能与图像分割后的其他区域有明显的区别, 故加入大小特征后分类效果能有略微的提高, 如图 3(b)所示。

4.2 分界面的建立

由于样本属于两个类, 一类为是头部区域, 一类为非头部区域。设 n 维训练样本为 x_1, x_2, \dots, x_N , 其中训练样本各个维度的物理意义在本应用中即为图像中已知头部区域大小, 长轴长度, 短轴长度, 周长, 形心, 颜色等特征。根据文献^[11], 其中最优化分类界面 H 为 $d(x) = \omega^T x - b = 0$ 。

给定训练样本集

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}, \quad (4)$$

其中 y_i 为某个区域是否为头部的标志位,取值为 1 或 0,1 表示是头部区域,0 表示非头部区域。

设分类界面为平面

$$\sum_{i=1}^N y_i \lambda_i x_i^T x + b = 0, 0 \leq \lambda_i, \quad (5)$$

通过训练所求的参数 $\lambda_i (i = 1, 2, \dots, N)$ 和 b 应满足

$$y_i (\sum_{i=1}^N y_i \lambda_i x_i^T x + b) \geq 1. \quad (6)$$

5 实验结果

图 4 为分割实验结果的对比,图 4(b)为简单快速的分水岭分割,由图可以看到,分水岭分割虽然简单、速度快,但会出现大量的过分割,会极大地影响后续决策分类的准确性和处理时间;而图 4(c)所示的基于 Graph Cut 的分割结果,虽然可以将目标与背景进行较好的分离,但并不能很好地将几个目标单独分割开来,同时由于 4 个目标人物衣着颜色与部分地板颜色相近,被误以为一个目标被合并至到一起,需要增加多个目标检测统计的环节,显得得不偿失。



(a)原始图像 (b) Watershed 分割结果
(a)Original image (b) Result of Watershed



(c) Graph Cut 分割结果 (d) Mean Shift 分割结果
(c) Result of Graph Cut (d) Result of Mean Shift

图 4 原始图像与分割后图像对比

Fig. 4 Comparison of original image and images after segmentation

图 4(d)为 Mean Shift 的图像分割结果,其中 $h_s=8, h_r=9$ 。该图像分割技术原理简单,计算速度较快,通常能在一次分割后形成大量小的模态区域。这样将直接将分析层次从像素域提升到特征域,从而使计算过程有更好的鲁棒性,且能更好的进行区域合并过程,最后以较少的处理时间作为代价得到一个较为准确的分割结果,满足系统对图像分割部分实时性和准确性的要求。

图 5 为原始图像与各种识别方法识别结果的对比。其中,(a)、(b)、(c)为原图。(d)、(e)、(f)为文献[5]中提到的基于头部轮廓椭圆检测的方法检测结果。在背景较为复杂时,经过边缘检测提取目标轮廓后,该方法后会受到背景以及目标服装配饰中形状类似物体的干扰,造成误检测,如(d)、(e)中的 1 号区域,实际为目标服饰的帽子,(f)中的 1 号区域,实际为背景中的消防栓,但由于轮廓形状均为椭圆,造成了检测误差。而基于彩色直方图的区域模板匹配的方法将某个区域的彩色直方图信息作为描述该区域的唯一信息,当某个区域(例如衣服的后帽,如图 5(g)、(h)、(i)中的 1 号区域以及黑色袖口下的人手,如图 5(g)、(h)、(i)中的 2 号区域)的色彩分布于人脸和头发的色彩分布比例相近时,会将此区域误判为人头区域,从而引入许多不可控的误差。本文提出的基于机器学习的检测方法与上文提到的这两种方法相比,首先将人头作为待识别的模式,在此应用中从根本上解决了摄像机特殊拍摄角度造成的困难,同时又加入颜色以及大小的特征,可以在某些特殊情况下避免帽子等干扰物带来的误差,能够较好的适应本应用的特殊性,如图 5(j)、(k)、(l)所示。

实验结果的平均准确率均能保持在 80% 以上,单次准确率最低不低于 70%,实际应用中,等待电梯的乘客有可能站立在镜头的边缘导致无法完全进入镜头、乘客完全背对镜头、脸部皮肤或头发无法暴露在图像中、光照突变等原因,均有可能造成检测误差。在图 5(i)中,1 号候梯乘客由于完全背对镜头,图像中采集不到其肤色信息,导致判决失败,形成漏检测。在图 6(b)中,1 号候梯乘

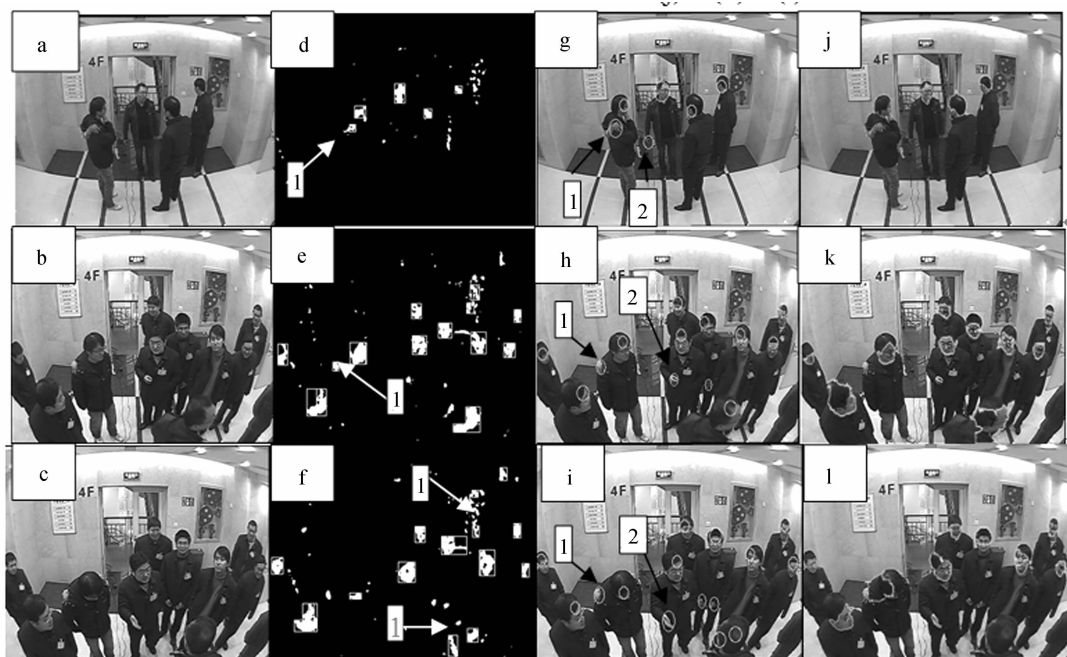


图 5 原始图像与识别结果对比

Fig. 5 Comparison of original image and images after recognition

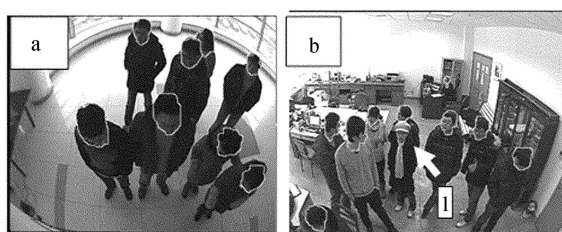


图 6 不同场景下的对比结果

Fig. 6 Comparison of images in other places

客由于头戴了一顶白色帽子,造成图像中采集不到其发色信息,导致判决失败,形成漏检测。

实验计算机环境配置为 Intel(R) Core(TM) 2 Duo CPU E7500, 1.96 GB, 测试系统环境为 Windows Xp, 总体检测的平均时间 418.69 ms, 标准差为 126.25 ms, 最大值为 773 ms, 最小值为 105 ms。实际 DSP 作为处理器的控制器中每次处理时间可以控制在 2 s 以内。

参考文献:

- [1] YANG G Z, HUANG T S. Human face detection in a complex background[J]. *Pattern Recognition*, 1994, 27(1):53-63.
- [2] 卢春雨, 张长水, 闻芳, 等. 基于区域特征的快速人

6 结 论

本文根据电梯群控系统的需求,提出了一种基于 Mean Shift 的图像分割与支持向量机(SVM)决策分类器为核心的视觉检测技术获得候梯人数的新方法。针对候梯人群图像采集角度、拍摄镜头的特殊性,通过分析计算选取了合适的摄像机,同时采用了 Mean Shift 的图像分割技术与 SVM 决策分类器相结合的方法对候梯人群头部进行了识别,快速地得到了较为准确的识别结果。实验数据表明,此检测方法准确率可保持在 80% 以上,能够满足电梯群控系统对候梯人群统计的准确性以及实时性的要求。该视觉检测系统有望用于电梯群控系统中。

脸检测法[J]. 清华大学学报:自然科学版,1999,39(1):101-105.

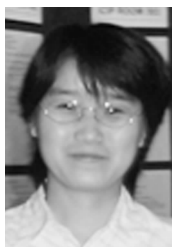
LU C Y, ZHANG C S, WEN F, *et al.*. A fast face detection algorithm based on region feature [J]. *Journal of Tsinghua University: Science and Technology*, 1999,39(1):53-63. (in Chinese)

- [3] HU Y W, ZHOU P, ZHOU H. A new fast and robust method based on head detection for people-flow counting system[J]. *International Journal of Information Engineering*, 2001,1:33-43.
- [4] ZIVKOVIC Z, KROSE B. An EM-like algorithm for color-histogram-based object tracking [C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004:798-803.
- [5] 于海滨. 基于头部特征提取的人体检测与跟踪及其应用[D]. 杭州:浙江大学, 2007.
YU H B. *Real-time human body detection and tracking based on head feature extraction and its application*[D]. Hangzhou: Zhejiang University, 2007. (in Chinese)
- [6] ZHANG E W, CHEN F. A fast and robust people counting method in video surveillance [C]. *International Conference on Computational Intelligence and Security*, 2007:339-343.
- [7] YURI B, VLADIMIR K. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision [J]. *IEEE Transactions on PAMI*, 2004, 26(9):1124-1137.
- [8] DORIN C. An algorithm for data-driven bandwidth selection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 24(5):281-288.
- [9] TAO W B, JIN H, ZHANG Y M. Color image segmentation based on Mean Shift and normalized cuts [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 2007,37(5):1382-1389.
- [10] SYLVAIN P, FREDO D. A topological approach to hierarchical segmentation using Mean Shift [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007:1-8.
- [11] 孙即祥. 现代模式识别[M]. 北京: 高等教育出版社, 2008.
SUN J X. *Modern Pattern Recognition*[M]. Beijing: Higher Education Press, 2008. (in Chinese)
- [12] SWAIN M J, BALLARD D H. Color Indexing[J]. *International Journal of Computer Vision*, 2002, 7(1):11-32.
- [13] CANNY J. A computational approach to edge detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8(6):679-698.
- [14] AKHIL K, KANCHAN W, Dr. AKHILESH U. Segmentation and counting of people through collaborative augmented environment [J]. *Global Journal of Computer Science and Technology*, 2011, 11(23):53-57.
- [15] QING Y. A robust method for counting people in complex indoor spaces [C]. *International Conference on Education Technology and Computer (ICETC)*, 2012:450-454.
- [16] HYUN H P, HYUNG G L, SEUNG-IN NOH, et al.. An area-based decision rule for people-counting systems [J]. *Multimedia Content Representation*, 2006, 4105:450-457.
- [17] DJAMEL M, KHEIR-EDDINE A, NICOLAS T. Fast people counting using head detection from skeleton graph [C]. *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2012:233-240.
- [18] KOIVUNEN T A. Noise-insensitive motion detector [J]. *IEEE TCE*, 1992,38(3):168-174.

作者简介:



张宇洋(1987—),男,浙江嘉善人,2010年于上海交通大学获得学士学位,主要从事计算机视觉检测方面的研究。E-mail: xingxing0468@hotmail.com



刘满华(1975—),女,江西吉安人,副教授,2006年博士毕业于南洋理工大学,主要从事图像处理、模式识别理论和应用研究。E-mail: mhliu@sjtu.edu.cn



韩 韬(1973—),男,山东莱芜人,2002年于上海交通大学获得博士学位,主要从事基于声表面波技术的射频标签和传感器方面的应用基础研究。E-mail: than@sjtu.edu.cn