

一种基于人头特征的人数统计方法研究

顾德军 伍铁军

(南京航空航天大学 机电学院 江苏 南京 210016)

摘 要: 对固定镜头下视频序列中人群的计数方法进行研究。根据人头的轮廓特性和颜色特性,提出一种人头检测的方法,较好地适用于稀疏人群和对遮挡不是十分严重的人群。把检测出的头部的中心点作为运动人体的特征点,利用基于目标运动连续性的匹配算法对入头部进行匹配跟踪,从而实现人数的计数。实验表明,该方法具有很好的实时性和较高的计数准确性。普通 PC 机即可实现。

关键词: 人数统计; 人头检测; 椭圆拟合; 最小二乘法; 人体跟踪计数

中图分类号: TP391.41 **文献标志码:** B **文章编号:** 1671-5276(2010)04-0134-05

Pedestrian Count Method Based on Head Feature

GU De-jun, WU Tie-jun

(College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: This paper discusses the pedestrian counting method in video sequences gathered from a fixed camera. According to the profile characteristic and color feather of a person's head, this paper proposes a novel algorithm which can be better used for sparse crowd to detect head. Then the central point of the head area is regarded as the feature point of a pedestrian. With the points, a match algorithm based on object motion continuity is used for head matching and tracking to implements a whole system for crowd counting. Large numbers of experimental results on several real video sequences demonstrates real-time effect and the high precision for various environments.

Key words: pedestrian counting; head detection; ellipse fitting; least square method; human tracking counting

0 引言

随着社会的不断发展,人数统计有着越来越重要的意义。传统的人数统计方法是人工计数或人工电子设备触发计数,不仅浪费人力物力而且效率低下。随着计算机视觉和图像处理技术的飞速发展,使得其在人数统计方面的应用成为可能。近年来关于人数统计的研究取得了很大的进展:文献[1]是一个可以在室外对人体进行实时检测和跟踪的视觉监视系统,它将外形分析与跟踪技术相结合来跟踪人体各个部分的位置并为人体的外形建立模型。文献[2-3]提出了一种基于颜色直方图的匹配跟踪算法,该算法在目标稀疏时效果很好,但是它没有考虑到目标间的遮挡以及目标各部分产生的自遮挡现象,另外该算法针对的是图像的彩色颜色空间,因此计算量较大。文献[4]采用立体摄像机,通过左、右眼获得的时空图像来实现人数统计。文献[5]采用立体摄像机,只对某一块感兴趣的区域进行检测,用高斯混合模型表示,跟踪个体的状态。并对情况进行分类,决定是否计数。以上方法要求采样频率较大,处理速度要求较高,其训练和分值的制定十分复杂,使得制定合适的阈值比较困难。综合以上方法的不足,提出一种基于人头检测的人数统计方法,该方法不仅计数准确,而且能满足较高实时性的要求,易于实现,采

用普通 PC 机即可处理图像数据。

1 运动目标分割

本文采用基于高斯统计模型的方法分割运动目标。这样不仅对如光照和外来无关事件的干扰有较好的抑制,而且能够提供最完全的轮廓数据。在视频序列图像中,每个像素点的灰度值是符合高斯分布的,因此可以计算:

$$\mu_n(x, y) = \frac{1-n}{n} \mu_{n-1}(x, y) + \frac{1}{n} f(x, y) \quad (1)$$

$$\sigma_n^2 = \frac{\sum_{i=1}^n (f_i(x, y) - \mu_n(x, y))^2}{n-1} \quad (2)$$

$\mu_n(x, y)$ 和 σ_n^2 分别表示前 n 帧图像在 (x, y) 点的均值和均方差。因此,利用有限帧图像可以建立一个背景模型:

$$B_0(x, y) = \frac{1}{m} \sum_{k=1}^m f_{ik}(x, y) \quad (3)$$

$f_{ik}(x, y)$ 满足 $|f_{ik}(x, y) - \mu_n(x, y)| \leq \beta \sigma_n(x, y)$, β 是一个预先设置的常量,可根据用来建模的视频帧数确定,当帧数较多时,可设置为 1。在获取运动场景过程中,由于背景的扰动、光照变化等背景的改变,背景模型必须进行及时更新。本文所采用的更新策略是,首先利用当前帧

作者简介:顾德军(1982—),男,江苏南通人,硕士研究生,研究领域为图像处理和模式识别。

与前一帧进行帧差,当两帧相差后某像素点灰度值大于某个阈值时,就判断为前景点,否则为背景点,对检测出运动前景区域点背景保留,而对被判断为背景的区域采用下面的公式进行更新:

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

(4)

$B_k(x,y)$ 为第 k 帧时刻的背景模型,其中 $0 \leq \alpha \leq 1$ 为更新速率,表示背景模型的更新快慢。使用更新的背景模型 $B_k(x,y)$ 对当前帧进行背景差分,大于某一阈值 T 的图像即为运动目标的点。现采用 OTSU 方法求取阈值 T 。图 1 为背景图像,图 2 为当前图像,图 3 为运动目标分割结果。



图 1 背景图像



图 2 当前帧图像



图 3 运动检测结果

从图 3 可以看出运动目标分割后的图像不仅存在噪声前景点,而且头部出现毛刺现象,这会对后期的工作带

来很大的影响。采用 Unger^[6]平滑再连通区域检测的方法去除毛刺和噪声。

Unger 平滑方法的思想是检查 3×3 窗口内图像结构的细节,而不是简单的统一取值为 1 的像素。 3×3 窗口如下图所示:其中 g 为当前点, $P_0 \sim P_7$ 为其 8 个邻接点,Unger 平滑准则为:当 $g=0$ 即 g 为白像素时,如果:1) P_0, P_1, P_2 中至少有 1 个等于 0,同时 P_4, P_5, P_6 中至少有 1 个也等于 0,或者 2) P_2, P_3, P_4 中至少有一个等于 0,同时 P_6, P_7, P_0 中至少有一个也等于 0,则 g 不变;否则 $g=1$ 。当 $g=1$ 即为黑像素时,如果 P_0, P_2, P_4, P_6 中至少有 3 个元素等于 0,则 g 改为 0;否则 g 不变。处理结果如图 4 所示。与之前的形态学运算去毛刺相比,不仅最大限度地保留了轮廓形状,而且速度也得到了大大提升。

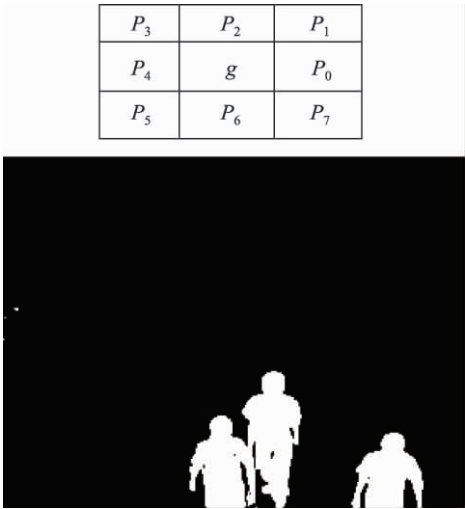


图 4 unger 处理后的图像

连通区域检测去噪即应用连通区域检测来找到属于同一个连通区域的所有像素点,并给予属于不同连通区域的像素点不同的标识,然后把面积小于预设阈值前景区域去除,留下来的前景区域就是运动对象。根据 CCD 固定的位置和运动目标之间的距离,本文的阈值选取为 300 个像素。图 5 为连通区域检测后的图像(以下简称图像)。从图中可以看出经过 unger 和联通区域检测后,图像中的毛刺和小块的噪声区域均被去除。



图 5 连通区域检测后图像

2 人头检测

2.1 特征点选取

在图像中,人头顶部点的特征为:该点在竖直方向局部最高处。即在图像中,该点的左边 k 个像素均为背景像素,并且该点上面一个点左右各 m 个像素及图像上边缘到这 m 个像素之间的像素也为背景像素。同时由于人的头发为黑色,所以在灰度序列图像中该点灰度值小于 T 。本文中 k 和 m 均取 3, T 取 40。求取人头部特征点的步骤为:

- 1) 提取图像中运动目标的轮廓。
- 2) 从图像左上角开始扫描图像,遇到像素值为 255 (运动目标轮廓点) 则检测灰度图中该点的像素值,若小于 40,则检测在图像中检测该点的左边 3 个点的像素值大小。否则,该点不可能为特征点,检测图像中下一点。
- 3) 若这 3 个像素值均为 0 (背景),则检查该点上面一个点和它左右各 3 个像素及图像上边缘到这 m 个像素之间的像素值的大小。否则该点不可能为特征点,舍弃,检测下一点。
- 4) 若再次检测的点均为黑色 (背景像素),则该点可能为人头顶部的点,存入动态数组中。否则该点不是人头部点,舍弃,检测下一点。

2.2 采样

征点选取完成之后,依次在特征点左右两边的轮廓线上采取 k 个数据点。由于采样单目摄像机,所以场景中的头部目标在图像不同位置大小会不同,采集的数据点的数目也会不同。所以首先对图像每一部分的人头大小进行事先估计,可以通过采集现场数据进行头部的大小估计。本文将图像划分成 3 个区域如图 6 所示,根据特征点所在的区域确定 k 的大小。

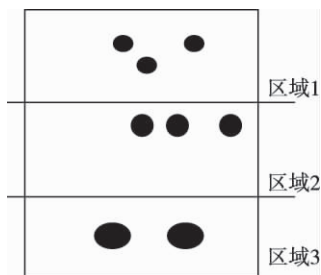


图 6 头部大小划分图

经过 *unger* 处理去除毛刺后,从特征点开始每采集一个数据点,该点的 y 坐标一定大于等于上一点的坐标(图像左上方为原点)。图 7 为两部分重合人头部轮廓 AB 为求得两个特征点,若检测的点的 y 坐标比上一点的 y 坐标值小,则认为取值越过了 C 点,去除该点,结束采样。

采集完毕后,从 A 点开始采集的数据点为轮廓线 CD 之间的像素点,从 B 点开始采集的数据点为轮廓线 CE 之间的像素点,如图 7 所示。

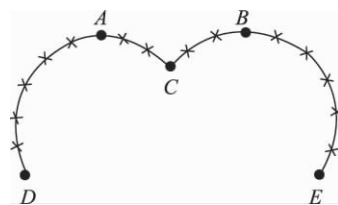


图7 两头部轮廓图

2.3 椭圆拟合

采集完数据点后 根据数据点做椭圆拟合。本文采用 Andrew Fitzgibbon 提出的直接最小二乘椭圆拟合方法^[7]。平面二次曲线的一般方程可以表示为:

$$F(\alpha, X) = \alpha^T X = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (5)$$

其中: $\alpha = [a \quad b \quad c \quad d \quad e \quad f]^T$

$X = [x^2 \quad xy \quad y^2 \quad x \quad y \quad 1]^T$ 。由几何知识可知, 当曲线方程系数满足 $b^2 - 4ac = -1$ 时, (5) 中表示的二次曲线方程为椭圆。如设:

$$C = \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

则最小二乘椭圆拟合可以转化成下面的问题:

$$\begin{cases} \alpha = \operatorname{argmin} \| D\alpha \| \\ \alpha^T C\alpha = 1 \end{cases} \quad (6)$$

其中: $X_i = [x_i^2 \quad x_i y_i \quad y_i^2 \quad x_i \quad y_i \quad 1]^T$;

$$\alpha = [a \quad b \quad c \quad d \quad e \quad f]^T;$$

$D = [X_1 \quad X_2 \cdots X_N]^T$ N 为参加拟合的数据点的个数。

对式(3)引入拉格朗日(Lagrange)系数并微分可得:

$$\begin{cases} S\alpha = \lambda C\alpha \\ \alpha^T C\alpha = 1 \end{cases} \quad (7)$$

式中 $S = D^T D$, 式(4)显然是广义特征矩阵求解问题, 所以根据广义特征矩阵求解方法可以方便求取, 式(3)的解为:

$$\alpha = \left(\sqrt{\frac{1}{u_i^T C u_i}} \right) u_i = \left(\sqrt{\frac{\lambda_i}{u_i^T S u_i}} \right) u_i \quad (8)$$

式中 λ_i 和 u_i 分别是式(7)的特征值和特征向量,可以证明式(6)解唯一。

该最小二乘椭圆拟合,采用的是广义特征根方法进行求解,算法简单高效,并且鲁棒性好。选用该方法对上述的采样点分别椭圆拟合。图8为利用该方法椭圆拟合的结果,设定椭圆长、短轴以及长短轴的范围,去除不符合条件的椭圆,结果如图9。当人头发生小部分遮挡时(图10),该方法也能检测出人的头部,检测结果如图11。

3 匹配跟踪计数算法

现把检测出的人体头部的中心点作为运动人体的特



图 8 椭圆拟合结果图



图 9 去除不符合条件椭圆后结果图



图 10 头部出现遮挡的情况

征点 这样在跟踪过程中若两人出现相互遮挡时 ,只要头部被检测出来 跟踪仍能被成功地执行 ,该方法的优点是实现简单 ,计算量少 ,并可以较好地解决人体遮挡问题。

3.1 基于目标运动连续性匹配算法

本文综合考虑了两帧间同一目标质心的 Euclidean 距离、运动速度和方向信息 给出了一种新的匹配算法。由于视频图像相邻帧间时间间隔很短 ,目标的某些运动参数并不发生突变 ,所以相邻两帧之间看做是匀速运动。对同一头部目标在两帧间其质心运动距离上限记为 th 。

设一图像序列 $I=(I_1\ I_2\ \cdots\ I_n\ \cdots)$ 第 n 帧图像 I_n 检



图 11 椭圆拟合结果图

测到 m 个目标 $s_{1\ n}\ s_{2\ n}\ \cdots\ s_{m\ n}$,目标 $s_{i\ n}\ (1\leq i\leq m)$ 的头部中心坐标 $(\ pos_x_i\ \ pos_y_i)$ 。第 $n-1$ 帧检测到的 k 个目标记作: $T=(\ T_1\ T_2\ \cdots\ T_k)$ 其 $T_i=(\ pri_x_i\ pri_y_i\ v_x\ v_y\ t\ label)$ $(1\leq i\leq k)$ 其中 $v_x\ v_y$ 表示第 i 个目标在第 $n-1$ 帧的速度 ,它由该目标点两次检测到的位置计算得到。 t 表示连续未检测到的帧数。 $label$ 为目标初始位置标志 ,表示目标进入检测区域前位置 ,在计数出检测线外标记为 1 ,在计数进检测线外记为 0(图 12) 。

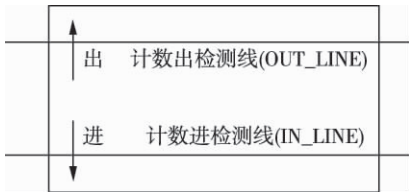


图 12 计数线设置

匹配算法如下:

1) 位置预测。由最后一次检测到的目标的情况预测该目标在当前帧的位置。以目标 T_j 为列 $(1\leq j\leq k)$, T_j 表示第 j 个目标最后一次被检测到的参数。预测其在当前帧的位置 $s_p(\ next_x_j\ next_y_j)$ 为:

$$next_x_j = pri_x_j + v_x \cdot t\ next_y_j = pri_y_j + v_y \cdot t$$

2) 计算 $s_{i\ n}$ 和 s_p 的 Euclidean 距离 ,得出测度值 $d = \| s_{i\ n} - s_p \|$ 。

3) 对目标链 T 中所有的目标 $T_j(j=1\ 2\ \cdots\ k)$ 计算目标的测度值 $D=(\ d_1\ d_2\ \cdots\ d_k)$ 。求得 $d_j = \min(D)$ 。若 $d_j < th$ 则匹配成功 ,判定 $s_{i\ n}$ 为 T_j 在第 n 帧出现的新位置 ,执行 4) ,否则 $s_{i\ n}$ 为一个新的目标 ,执行 5) 。

4) 更新 T_j 。 $T_j=(\ pos_x_i\ pos_y_i\ v'_x\ v'_y\ t\ label)$ pos_x_i 为 $s_{i\ n}$ 的中心坐标 v'_x 和 v'_y 由 $s_{i\ n}$ 中心和更新前 T_j 中心坐标计算得到。 t 设为 1。 $label$ 保持不变。

5) 在目标链尾部加入目标 T_{k+1} 。 $T_{k+1}=(\ pos_x_i\ pos_y_i\ v'_x\ v'_y\ t\ label)$ $pos_x_i\ pos_y_i$ 为 $s_{i\ n}$ 的中心坐标 , v'_x 和 v'_y 为新目标的初始速度 ,设为零。 t 设为 1。比较目标

中心点与计数进检测线和计数出检测线距离,若靠近计数出检测线则判定该目标从计数出检测线外进入检测区, label 设为 1,否则设为 0。

3.2 跟踪计数算法

由于目标可能被遮挡或者其他原因不能被检测出来,所以当目标没有被检测出来时,不能判断目标已经在检测区内消失。本文采用等待 n 帧来判断目标是否在检测区消失。跟踪计数算法如下:

- 1) 判断目标链 T 是否为空,若为空,则 $s_{1\ n}, s_{2\ n}, \dots, s_{m\ n}$ 均为新目标,执行匹配算法 5),再处理第 $n + 1$ 帧。若不为空,把目标链中所有目标标记为待匹配状态。
- 2) 将 $s_{1\ n}, s_{2\ n}, \dots, s_{m\ n}$ 按照基于目标运动连续性匹配算法与目标链中待匹配的目标匹配。若 T_j 与 $s_{i\ n}$ 匹配成功,将 T_j 标记为已检测状态。继续检测 $s_{i+1\ n}$,直到 $j = m$ 时结束。
- 3) 检测目标链 T ,若发现仍有目标 $T_j(1 \leq j \leq k)$ 处于待匹配状态,则执行等待 n 帧判断的方法:若 $t > n$,执行 4)。若 $t \leq n$,将 t 增加 1, T_j 标记为已检测状态,其余参数不变。
- 4) 比较 T_j 中目标中心点与计数进检测线和计数出检测线距离,若靠近计数出检测线,且 label 为 0,则出的人数增加 1,把 T_j 从目标链中删除。若靠近计数进检测线,且 label 为 1,则进的人数增加 1,把 T_j 从目标链中删除。否则,把 T_j 从目标链中删除,人数不变。

按照上述方法,把实际拍录的视频数据进行了测试,得出的结果如表 1 所示。

表 1 实际值与仿真测得值对比

	实际值		测量值	
	进入	出来	进入	出来
视频 1	3	0	3	0
视频 2	0	3	0	3
视频 3(1 人折回)	2	0	2	0
视频 4(2 人折回)	1	0	1	0
视频 5(3 人折回)	0	0	0	0

实验中一台普通的 PC 机(P4 2.8 CPU,512M 内存)完全可能达到实时性处理要求。表 1 中第一个视频 3 个人进入,中途没有折回现象。第二个视频 3 个人离开,中途也没有折回现象,这两种情况比较简单。在第三个视频中首先进入 3 个人,有一人中途折回,按照技术规则,不予计数。第 4 个视频录像中,有 2 个人在进入后折回,所以这 2 个人同样不能计算在进入人数之中。最后一个图像是 3 个人进入后,都折回的情况,所以没有人进入。实验结果表明该算法具有很好的实时性、鲁棒性及较高的计数准确性。

4 结语

本文提出了一种新的人数统计的方法。根据人的头部轮廓特性和颜色特性,提出一种人头检测的方法,不仅可以处理稀疏人群而且对遮挡不是十分严重的人群处理

效果也较好。在人数计数上,采用了最近邻匹配跟踪方法,利用检测出的一个目标和目标链中的所有目标进行匹配,这样有效克服了错误匹配,增加了简单的运动位置预测处理过程,计数的准确性得到了很大的提高。实验结果表明该方法具有很好的实时性、鲁棒性及较高的计数准确性,而且易于实现。

本方法仍有不足之处,检测区域内目标长时间处于被遮挡时不能准确计数,今后将考虑利用多摄像头从不同角度进行处理。

参考文献:

[1] Haritaoglu I, Harwood D, Davis L. W4: Real - time surveillance of people and their activities [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000, 22(8) : 809-830.

[2] Haritaoglu I, Flickner M. Detection and tracking of shopping groups in stores [C] // Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [J]. Kauai, HI, USA: IEEE Computer Society, 2001.

[3] Lu Wenmiao, Tan Yap - peng. A color histogram based people tracking system [C] // Proc. of IEEE International Symposium on Circuits and Systems. Sydney, Australia: [s. n.], 2001.

[4] Terada K, Yoshida D, Oe S, et al. A counting method of the number of passing people using astereo camera [J]. Industrial Electronics Society, 1999(3) : 1318-1323.

[5] Beymer D. Person counting using stereo [J]. Human Motion, 2000(8) : 127-133.

[6] Victor Wu, Raghavan Manmatha, Edward M, et al. TextFinder: An automatic system to detect and recognize text in images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999 21(11) : 1224-1229.

[7] Fitzgibbon A, Pilu M, Fisher R B. Direct least square fitting of ellipse [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(5) : 476-480.

[8] 邓玉春,姜昱明,张建荣. 视频序列图像中运动对象分割综述 [J]. 计算机应用研究, 2005, 4(1) : 8-11.

[9] Zhao T, Nevatia R. Tracking multiple humans in complex situations [J]. IEEE Trans. Pattern Anal. Mach. Intell, 2004: 26(9) : 1208-1221.

[10] 王成儒,顾广华. 一种采用背景统计技术的视频对象分割算法 [J]. 光电工程, 2004, 31(8) : 57-60.

[11] Andrew Gilbert, Richard Bowden. Incremental, scalable tracking of objects inter camera [J]. Computer Vision and Image Understanding, 2008(111) : 43-58.

[12] Black J, Ellis T. Multi camera image tracking [J]. Image Vis, Comput, 2006: 24(11) : 1256-1267.

[13] Byeoung-Su Kim, Gwang-Gook Lee, Ja-Young Yoon, et al. A method of counting pedestrians in crowed scenes. Proceedings of the 4th international conference on Intelligent [J]. Computing, 2008, 1117-1126.

收稿日期: 2009 - 11 - 26