

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2013)04-0392-07

论文引用格式: 覃勋辉, 王修飞, 周曦, 刘艳飞, 李远钱. 多种人群密度场景下的人群计数[J]. 中国图象图形学报, 2013, 18(4): 392-398.

多种人群密度场景下的人群计数

覃勋辉, 王修飞, 周曦, 刘艳飞, 李远钱

中国科学院重庆绿色智能技术研究院, 重庆 401122

摘要: 公共场合中采用摄像机实现人群计数在智能安防领域具有重要价值, 但摄像机透视效果、图像背景、行人相互遮挡等因素制约着人群计数研究的发展和应用。提出一种采用回归模型估计人数的算法。首先, 为了消除摄像机透视对图像特征的影响, 用图像中行人身高作为尺度基准将图像分成多个子图像块。其次, 采用 simile 分类器优化子图像块的先进局部二值模式(ALBP) 纹理特征, 并根据子图像块的人群密度, 采用两种核函数的支持向量回归机(SVR) 建立输入特征和子图像块人数的关系。最后, 相加所有子图像块人数得出图像人数。实验结果表明, 本文算法测试稀疏人群的绝对误差约为 1 人, 测试拥挤人群的相对误差小于 10%, 是一种准确率高适用性强的人群计数算法。

关键词: 人群计数; simile 分类器; 支持向量回归机; 人群密度估计

Counting people in various crowd density scenes using support vector regression

Qin Xunhui, Wang Xiufei, Zhou Xi, Liu Yanfei, Li Yuanqian

Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 401122, China

Abstract: The use of video surveillance in for people counting public places has an important value in the field of intelligent security. However, there are several factors such as camera perspective, background clutter, and occlusions, which restrict its development and application of the study. An algorithm based on the regression model is proposed for estimating the number of people. First, in order to eliminate the effect of the camera perspective on the image features, the input image is divided into several sub-image blocks according to the change of pedestrian height in the image. Second, the simile classifier is used to improve the advanced local binary patterns (ALBP) texture feature of the blocks. Then, according to the crowd density, we use the support vector regression (SVR), which has two kernel functions to establish the relationship between input features and the number of people. Finally, adding the number of persons of all sub-image blocks gives us the total number of people on the image. Experimental results show that the absolute error of the sparse population is approximately one person using the presented algorithm and the relative error of the testing crowded population is less than 10%. This therefore demonstrates the high accuracy of this algorithm, which can be applied for people counting in video surveillance.

Key words: counting people; simile classifier; support vector regression; crowd density estimation.

收稿日期: 2012-08-08; 修回日期: 2012-10-25

基金项目: 重庆市重大科技攻关项目(cstc2011ggC40009)

第一作者简介: 覃勋辉(1986—) 男, 研究实习员, 2012 年于西北工业大学获控制科学与工程硕士学位, 主要研究方向为视觉导航、大规模人群计数、人脸识别等。E-mail: qinxunhui@cigit.ac.cn

0 引言

公共场合中通过摄像机实现人数计数具有重要的研究价值。比如:候车大厅中人群计数的结果,可优化公共交通的调度;某区域中人数的急剧变化可能会导致意外事件的发生,又可能是意外事件发生的结果。因此,人群计数是计算机视觉和智能视频监控领域的重要研究内容。

目前主要有两种人群计数的方法:1)直接法^[1-3]:在人群较稀疏的场景中,通过检测视频中的每一个行人,进而得到人群计数的结果。2)间接法^[4-12]:建立图像特征和图像人数的回归模型,通过测量图像特征从而估计场景中的人数。由于拥挤情况下采用直接法容易受到遮挡等难点问题的影响,而间接法从人群的整体特征出发,具有大规模人群计数的能力,所以本文选用间接法。

图像特征、摄像机透视效果是影响间接法人群计数效果的主要因素。间接法常用的特征有:像素统计特征(运动图像块的大小^[4-6]、运动特征点个数^[7-8])、纹理特征^[9](GLDM)、梯度统计特征^[10-12](HOG)等。但是,由于遮挡和透视效果的影响,现有图像特征难以达到较好的效果。另外,摄像机透视效果导致不同位置的行人在图像平面呈现不同的形状和大小,所以难以建立图像特征和人数关系的模型。文献[9,11-12]通过摄像机标定方法求解摄像机的透视关系,再利用此关系归一化图像特征,取得了一定的效果。但是,摄像机标定过程限制了该方法的推广。另外,当人群较拥挤时,由于文献[8,10]算法难以分割出运动物体,导致算法失效。

提出了一种能够在多种人群密度场景下实现人群计数的算法。本文算法主要有以下创新:1)用图像中行人身高作为尺度基准计算透视关系,并将图像分成若干个具有不同透视程度的图像子块,图像总人数是所有图像子块中图像人数之和;2)采用simile分类器^[13]优化图像特征:选择具有代表性的“参考图像”,计算输入图像特征和“参考图像”的相似性,该相似性组成的向量作为新的特征;3)根据不同人群密度采用不同的支持向量回归机(SVR)回归模型估计人数。

1 图像分块

图像分块是一种离散化透视效果的方法。摄像

机透视效果对人群计数问题的影响主要表现在:1)造成位于不同景深的行人在图像平面具有不同的形状和大小,目前还没有特征描述行人时同时具有形状和尺度不变性的能力;2)造成远离摄像机区域的人群更密集,相互遮挡更严重,增加了人群计数问题的难度。

因此,求解透视关系是提高间接法人群计数算法性能的关键步骤^[12],目前常用摄像机标定方法计算透视关系:1)已知摄像机的内外参数,根据摄像机投影矩阵计算透视关系;2)在地面上指定4个点的相对位置关系(多为指定矩形^[11]),计算图像平面到摄像机平面的单应矩阵。文献[9,11-12]通过摄像机标定方法求解摄像机的透视关系,再利用此关系归一化图像特征,都取得了较好的人群计数结果。但是,这些算法需要在摄像机安装现场标定摄像机,阻碍了算法的应用推广。

解决人群计数问题也可以未知摄像机参数的具体数值,但需要计算行人姿态或者大小在图像中变化的规律。在图像左上坐标系中,行人身高(或者行人面积)随着图像纵坐标减小而增加,故该摄像机透视关系可以表示为图像中行人身高和图像位置的函数。基于该函数关系将连续的透视效果离散化,将图像分成多个子图像块,并认为每个子图像块中的透视程度一致,以消除透视效果对特征的影响。具体方法如下:

假设摄像机拍摄平面场景,所有行人的身高均为 h ,图像中 y 轴方向上行人的身高为 H ,图像中行人脚的纵坐标为 y_1 ,那么有数学关系^[6]

$$H = ky_1 + b \quad (1)$$

式中 k 、 b 是反映图像平面上行人大小与图像位置关系的两个变量。对于待估计人数的视频或者图像,用户仅需在图像中指定多个行人的头和脚的位置便可求解式(1),根据式(1)算法自动将图像分成多个图像子块。值得注意的是,为了得到较准确的 k 、 b 值和较好的分块效果,最好在多幅图像的不同位置指定多个行人的位置。

图1是按照上述方法计算的图像分块结果。可以看出,每个子图像块中行人的大小并没有明显的差异,便于用同一个特征描述行人,算法通过估计每个图像子块人数以达到人群计数的目的。对于只有部分身体在子图像块中时,按照行人的面积计数。



图 1 图像分块结果

Fig. 1 The result of image cells

2 特征提取

为了消除图像背景对人群计数算法的影响,大多数算法^[4-6,10-12]都会先去除图片背景,再提取边缘、面积等特征描述行人。显然,去除背景有利于更直接地描述和提取行人的特征。但是,由于光照变化、行人拥挤程度、背景颜色等多种因素干扰,都使得去除图像背景成为一项较难的工作。图像分块后,消除了摄像机透视的干扰,图像统计特征由背景特征和行人特征组成,此时特征且具有了区分背景和行人的能力,故本文算法直接在原始图像上提取特征。

2.1 图像 ALBP 特征

局部二值模式(LBP)特征是一种具有光照不变性的局部纹理描述特性^[14],它通过比较中心像素点和其邻域像素点的值进而产生一串比特流来描述该点的纹理特征。一致局部二值模式(ULBP)特征改进了原始的LBP特征,它不仅将原始的LBP特征从128维降低到59维,降低了特征提取的复杂度,同时还增强了该特征抵抗噪声的能力。先进局部二值模式(ALBP)特征降低了ULBP特征的维数,从原始的59维ULBP特征降低为16维特征。ULBP特征的编码可看做是“1”弧线(白弧)和“0”弧线(黑弧)连接而成。用白弧长度和白弧起点表示ULBP特征,即得到ALBP特征^[15]。

2.2 simile 分类器优化特征

simile分类器是文献[13]提出的一种优化特征的方法,且被证实该方法在人脸识别问题中显著的提高了识别率。simile分类器的基本思想是:物体

的属性可以用与它同范畴的其他物体的属性比较得到。对于人群计数问题,一幅图像的人数可以和其他图像的人数相比较得到。虽然2.1节中已经叙述了ALBP特征,但是该特征统计了背景纹理和行人纹理,缺乏准确描述行人的能力,故采用simile分类器增强ALBP特征描述人数的能力。具体算法如下:

算法功能:采用simile分类器优化ALBP特征。

算法输入:测试图像 I 。

算法输出:子图像块对应的 k 维图像特征 F 。

算法流程:

1) 计算灰度图像对应的LBP特征图;

2) 基于LBP特征图统计16维的ALBP特征,并将该特征归一化;

3) 生成“参考图像”,选择 K 幅同测试图像具有相同位置的训练图像,按照步骤1) 2) 计算它们的ALBP特征,再采用k-means聚类得到 $k(k < K)$ 个聚类中心,即为“参考图像”;

4) 对于测试图像 I ,按照步骤1) 2) 计算它的ALBP特征 f_{test} ,并计算该特征与第 i 个“参考图像” f_{ref} 的相似性 $s_i = \sum_{j=1}^{16} f_{\text{test}}(j) \cdot f_{\text{ref}}(i, j)$,再将 k 个相似性组合为一个新的特征 $f_{\text{new}} = [s_1 \ s_2 \ \cdots \ s_k]$;

5) 归一化后的 $F = [s_1 \ s_2 \ \cdots \ s_k] / \sum_{i=1}^k (s_i^2)$ 作为子图像 I 的特征。

图2是从视频图像序列中截取的3帧图像。可以看出,图2(a)(b)相似,图2(a)(c)不相似。在图像特征空间中,基于ALBP特征的子图2(a)(b)相似度为0.99,高于子图2(a)(c)的相似度0.82,则推测图2(a)人数更接近于图2(b),推测

结果和实际相符,说明 simile 优化特征方法可行。



图2 simile 分类器分析图像人数

Fig. 2 Analysis of number of people using simile classifier

3 建立回归模型

用支持向量回归机建立图像特征和图像人数的数学关系,该回归机具有结构简单、全局最优解、泛化能力强等优势。对于图像子块 i ,首先根据训练样本中图像特征和图像人数训练 SVR 模型。然后,该 SVR 模型根据输入图像特征便可估计图像人数。采用 SVR 模型估计人数的表达式为

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) \cdot K(\bar{x}_i, x) + b \quad (2)$$

式中 a_i 和 a_i^* 是拉格朗日乘子, \bar{x}_i 是支持向量, x 是图像特征, $f(x)$ 是估计的图像人数, $K(\bar{x}_i, x)$ 代表核函数, b 是常数。当人群稀疏时,通常认为图像统计特征和图像人数是线性关系^[12],所以设 $K(\bar{x}_i, x)$ 为线性核。但是,当人群密度较高时由于行人之间相互遮挡,线性核函数已经不再适用于人群计数。另外,SVR 模型中训练集人数分布范围也会一定程度

影响模型估计值。如果用人数偏低的样本训练模型,那么该模型的估计人数值很有可能偏低。所以,根据不同的人群密度,需要使用不同的核函数 $K(\bar{x}_i, x)$ 才能够得到更满意的实验结果。

在人群密度估计问题中,文献[15-16]设定行人面积占图像块面积的60%以上时为图像块人群拥挤状态。基于该算法估计人群密度已达到96%的正确率,因此首先用该人群密度估计算法判断图像子块中的人群是否拥挤,然后根据判断结果采用两个不同的 SVR 核函数。

人群拥挤时采用径向基核函数为

$$K(\bar{x}_i, x) = e^{-g \cdot (\bar{x}_i - x)^2} \quad (3)$$

人群稀疏时采用线性核函数

$$K(\bar{x}_i, x) = \bar{x}_i^T \cdot x \quad (4)$$

4 实验及其分析

使用4个具有不同人群密度场景(0~91人)的视频测试算法性能,这些视频的样本图像及其图像分块方式如图3所示。场景1和场景2人群稀疏,场景3、场景4人群密集。4个场景都拥有一定程度的背景图案。场景3和场景4由于行人之间相互遮挡严重,采用行人检测的方法难以取得较好的行人计数效果。同时,对于场景3和场景4中的部分区域人群密度过高的情况,基于去除背景的人群计数算法也难以取得较好效果。



图3 测试视频的图像样本

Fig. 3 Image samples of the test videos

为了评估算法的性能,设置平均绝对误差(MAE)和平均相对误差(MRE),即

$$MAE = \frac{1}{N} \sum_{i=1}^N |P(i) - T(i)| \quad (5)$$

$$MRE = \frac{\sum_{i=1}^N |P(i) - T(i)|}{\sum_{i=1}^N T(i)} \times 100\% \quad (6)$$

式中 N 是测试视频的帧长度; $P(i)$ 是算法估计图像 i 的人数,它是所有图像子块估计的人数之和; $T(i)$ 是由人为标定的图像 i 的人数,作为参考值。

4.1 算法性能

从场景1—4视频中各选取300幅图像,其中100幅图像作为训练样本,剩余的200幅作为测试样本。本文算法得到如表1所示的实验结果。

表 1 算法性能

Table 1 Performance of the proposed algorithm

视频 场景	训练集		测试集		视频人数		
	MAE	MRE/%	MAE	MRE/%	最小	最多	平均
场景 1	0.4	10.3	0.7	15.3	0.0	8.7	4.4
场景 2	0.6	9.1	1.1	14.4	0.0	13	6.9
场景 3	2.8	4.5	5.1	8.0	42	82.0	63.5
场景 4	2.5	3.5	3.9	5.4	55.7	91.5	71.3

由表 1 的实验结果可知,本文算法测试稀疏人群的绝对误差约为 1 人;测试拥挤人群的相对误差小于 10%。本文算法在人群稀疏场景和人群密集场景都适用,且并不需要完成摄像机标定工作和去掉背景工作,是一种准确率高适用性强的人群计数算法。

4.2 与文献[1]算法比较

本文算法较传统的直接法^[1]具有更高的准确性。图 4 和图 5 分别是本文算法和文献[1]算法对场景 3 第 127 帧的人群计数结果。



图 4 本文算法测试场景 3 第 127 帧的结果
Fig. 4 Counting results of 127 frame in scene 3



图 5 文献[1]算法测试场景 3 第 127 帧的结果
Fig. 5 Counting results of 127 frame in scene 3

本文算法中图像人数由子图像块的估计人数相加得到。图 4 中,每个子图像块左上方“c”值为人数标定值,“p”值为算法估计值。图 4 右下角“Tcab”代表标定的图像人数为 73.5 人,“Tpre”代表估计的图像人数为 71.0 人,它是所有子图像块的估计人数总和。可以看出,本文算法估计子图像块人数和图像总人数都较为准确。

文献[1]算法通过检测行人头肩以达到人群计数的目的。图 5 为该算法测试人群密集时的实验结果,蓝框为检测到的头肩。可知,该算法共检测头肩 54 个(参考值 73.5),该结果包括了误检的 10 多个头肩,同时还漏检了 20 多个头肩。当人群密集时,由于行人间的相互拥挤遮挡造成图像中行人头肩不完整和行人姿态各异,故该算法难以取得较理想的实验效果。

4.3 不同特征的比较

为了证明用 simile 分类器优化图像 ALBP 特征的有效性,表 2 对比了本文特征 SALBP(ALBP + simile) 和其他特征估计 4 个场景人数的性能。

表 2 采用不同特征时的算法性能

Table 2 The algorithm performance using different features

特征	GLDM	HOG	ALBP	SGLDM	SHOG	SALBP
维数	16	128	16	30	30	30
MAE	5.7	3.0	2.9	3.5	2.9	2.7
MRE/%	16.3	8.6	8.2	10.1	8.3	7.7

表 2 中,SGLDM 和 SHOG 代表用 simile 分类器优化后的图像 GLDM 和 HOG 特征。从表 2 实验结果可知:1) 单独使用 ALBP 特征估计人数的实验效果要优于单独使用 HOG 和 GLDM,因为 ALBP 特征是具有低特征维数和抗细小噪声的优点;2) 使用 simile 分类器优化特征后的实验效果要好于未优化的实验效果,尤其是特征本身效果不理想时,使用 simile 分类器优化特征的效果提升更明显。

4.4 不同的核函数的比较

本文算法指出用线性核函数和径向基核函数分别估计图片人群稀疏和人群拥挤时的人数。为了表明本文算法的有效性,实验比较本文算法和两种单一核函数估计场景 3 的实验效果,如图 6 所示。

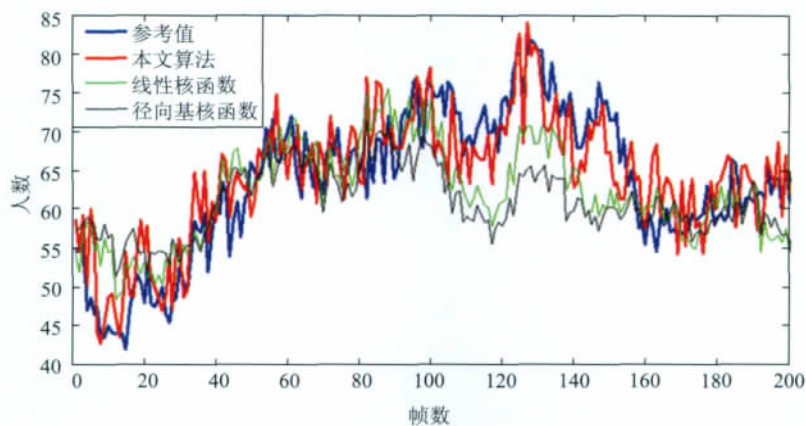


图6 采用不同核函数估计场景3人数

Fig. 6 People number estimation in scene 3 adopting different kernel functions

从图6可以看出,本文算法估计的人数曲线在参考值上下小范围内波动,而采用单一的核函数方法估计人数密集高的图像时(90~150帧),估计值明显地小于参考值,故采用两种核函数组合的方法效果优于单一的核函数。这是由于采用单一核函数有以下不足:1)训练模型容易受训练集样本人数分布的影响;2)人群密度较高时,行人容易发生相互遮挡,改变了人数稀疏时人数和特征的线性关系。

5 结 论

提出一种基于支持向量回归机的人群计数算法。使用图像中行人高度作为尺度基准将图像分为多个子图像块,再通过向SVR模型输入图像特征以估计子图像块人数,所有子图像块人数之和即为图像人数。通过实验和理论分析,本文算法具有以下主要特点:1)算法适用性强,算法不需要传统摄像机标定工作计算图像透视关系,也不需要背景建模去掉图像背景;2)算法准确性高,在图像人群稀疏时估计的平均绝对误差约为1人,在图像人群密集时估计的平均相对误差小于10%,这是因为不仅采用了simile分类器优化了图像ALBP特征,而且根据不同的人群密度采用不同的SVR回归模型估计人数。

虽然在大规模人群计数问题上取得了一些进展,但由于受图像背景和透视关系的影响,同此问题的其他研究成果一样,还不能在训练场景和测试场景不同时取得较好效果,这将是下一步研究

的重点。

参考文献 (References)

- [1] Li M, Zhang Z X, Huang K Q. Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection[C]// Proceedings of the 19th International Conference on Pattern Recognition. Florida, USA: IEEE, 2008: 1-4.
- [2] Wu B, Nevatia R. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors[C]// Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005: 90-97.
- [3] Zhao T, Nevatia R, Wu B. Segmentation and tracking of multiple humans in crowded environments[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(7): 1198-1211.
- [4] Choudri S, Ferryman J M, Badii A. Robust background model for pixel based people counting using a single uncalibrated camera[C]// Proceedings of the 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Snowbird, Utah: IEEE, 2009: 1-8.
- [5] Hou Y L, Pang G K. People counting and human detection in a challenging situation[J]. IEEE Transactions on Systems Man and Cybernetics, 2011, 41(1): 24-33.
- [6] Celik H, Hanjalic A, Hendriks E A. Towards a robust solution to people counting[C]// Proceedings of IEEE International Conference on Image Processing. Atlanta, USA: IEEE, 2006: 2401-2404.
- [7] Conte D, Foggia P, Percannella G. A method for counting people in crowded scenes[C]// Proceedings of the Seventh IEEE International Conference on Advanced Video and Signal based Surveillance. Klagenfurt, Austria: IEEE, 2011: 111-118.

- [8] Conte D , Foggia P , Percannella G. Counting moving people in videos by salient points detection [C]// Proceedings of the 20th International Conference on Pattern Recognition. Istanbul , Turkey: IEEE , 2010: 1743-1746.
- [9] Wu X Y , Liang G Y , Lee K K. Crowd density estimation using texture analysis and learning [C]// Proceedings of the IEEE International Conference on Robotics and Biomimetics. Kunming , China : IEEE 2006: 214-219.
- [10] Chan A B , Liang Z S , Vasconcelos N. Privacy preserving crowd monitoring counting people without people models or tracking [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Flofida , USA: IEEE , 2008: 1-7.
- [11] Chan A B , Vasconcelos N. Counting people with low-level features and bayesian regression [J]. IEEE Transactions on Image Processing , 2012 21(4) : 2160-2177.
- [12] Kong D , Gray D , Tao H. A viewpoint invariant approach for crowd counting [C]// Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong , China: IEEE , 2006: 1187-1190.
- [13] Kumar N , Berg A C , Belhumeur P N. Attribute and simile classifiers for face verification [C]// Proceedings of IEEE International Conference on Computer Vision. Kyoto , Japan: IEEE , 2009: 365-372.
- [14] Maenpaa T , Pietikainen M. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence , 2002 , 24(7) : 971-987.
- [15] Ma W H , Huang L , Liu C P. Advanced local binary pattern descriptors for crowd estimation [C]// Proceedings of Pacific-Asia Workshop on Computational Intelligence and Industrial Application. Wuhan , China: IEEE , 2008: 958-962.
- [16] Ma W H , Huang L , Liu C P. Crowd density analysis using cooccurrence texture features [C]// Proceedings of the 5th IEEE Conference on Computer Sciences and Convergence Information Technology. Seoul , Korea: IEEE , 2010: 170-175.