文章编号:1671-4598(2017)05-0076-04 DOI:10.16526/j.cnki.11—4762/tp.2017.05.021 中图分类号:TP391.4 文献标识码:A

基于多尺度协同的人头检测方法

彭景维,童基均
(浙江理工大学 信息学院,杭州 310018)

摘要:针对HOG特征本身不具有尺度不变性,在实际应用中仅能检测出与样本图片大小相差不大的目标对象这一弊端,提出多尺度窗口融合的头部检测的方法;利用线性支持向量机在分类决策方面的优势,与提取的HOG特征结合作分类器的离线训练;在实时的目标检测阶段,采用高斯金字塔式缩放对输入的视频序列作多尺度处理,得到对应的不同分辨率下的待检测帧,在不同的尺度空间作人头的扫描检测并存储结果;之后融合各尺度的检测结果并在相应位置决策标定;实验对某监控视频作检测分析,结果表明,该方法在检出率、召回率、准确度等方面均有较大提升。

关键词:多尺度;金字塔变换;窗口融合;梯度直方图

Head Detection Method Based on Multi—scale Collaboration

Peng Jingwei, Tong Jijun
(School of Information Science and Technology, Zhejiang Sci—Tech University, Hangzhou 310018, China)

Abstract: The HOG feature was not scale invariance and could only detect the targets which had similar size with sample image in practical application, proposed a method of head detection based on multi—scale windows fusion. The HOG features were extracted and the support vector machinewas used as classifier. In real—time detection of moving targets, the Gaussian Pyramid was used to make multi—scale decomposition for a sequence of input video frames, and got frames of different resolutions, and then run head detection at different scales and storage result. To improve the detection accuracy and efficiency, all detection results of each scale space were fused and got their corresponding locations signs. One monitoring video was tested, and the experiment results showed that the proposed method could improve the detection rate, recall and detection accuracy.

Keywords: multi—scale; pyramid transform; window fusion; gradient histogram

0 引言

计算机智能视频监控系统成为近年来的热点研究方向,行人人头检测技术在视频监测领域的应用也越来越普及,比如安防、人流量统计分析、舞台虚拟编排的走位校正等应用正发挥着越来越大的社会效益和利用价值。人头检测的研究发展至今,研究人员已做了大量的探索,Zhou等[1]对于较为拥挤的人群场景,利用多核学习技术并结合梯度直方图和局部二元模型的特征集训练一个头部检测的分类器,并用该技术建立模型框架,实现头部的检测及形状准确识别。在Guan等[2]的系统中,忽略人体其它部位的形态,将头部建模为椭圆形,并与基于颜色直方图的特征融合,用融合后的特征为检测子,实现人的头部的验证和追踪。Aziz等[3]提出了一种基于骨架图形的人头检测方法,该方法利用一种适应骨架图形分析技术,在拥挤环境中辨别每个选定斑点的轮廓信息,以达到检测行人头部的目的。

方法均可检测出视频中的行人,然而实时视频中的准确度、鲁棒性等均有待提高,如文献[1]方法在场景复杂时漏检较高,文献[3]方法在视频中人数不多的时候较为有效。本文提出一种检测实时场景中的人头部的方法,将Dalal等[4]提出的梯度直方图(Histogram of Gradient, HOG)特征运用到人头检测上,针对传统HOG本身不具有尺度不变性等弊端,提出一种多尺度协同及检测窗口融合的方法,以提高检测精度和效率。

收稿日期:2016-12-23; **修回日期**:2017-01-05。

基金项目:浙江省重点研发计划(2015C03023);浙江理工大学"521人才培养计划"。

作者简介:彭景维(1989-),男,湖北孝感人,硕士研究生,主要从事计算机视觉、图像处理方向的研究。

1 多尺度协同分析

传统HOG特征本身不具备尺度不变性,检测窗口大小固定,对于实时视频场景中的位置、大小等不断变化的人头目标很难及时作出有效响应,本文针对这一弊端作出改进,主要按照以下两个模块进行:一是分类器的离线训练阶段,主要表现在正负样本的采集、HOG特征的提取过程、SVM分类学习并生成头部分类器等几个方面;二是实时的在线目标检测阶段,主要是图像尺度缩放,并对缩放后的图像提取HOG特征,这样减少滑动窗口分类时的重复提取,接着做检测窗口的密集扫描处理,并用离线训练的分类器作检测窗口的分类,保存分类结果,高斯金字塔的融合处理及标定最终结果等过程。本文所述的检测方法基本流程如图1所示。

1.1 分类器离线训练

样本库的建立:

对于行人的头部检测,样本质与量的好坏对于检测结果准确与否扮演重要角色。分类器离线训练过程的第一步是选取样本,构建训练分类器模型所需的样本库,包括正样本(头部样本)和负样本(非头部样本)的选择。

1)头部样本选取头部样本作为训练数据集的正样本,其选取的质与量对于建立的分类器检测性能的好坏和检测结果的准确性有直接的影响。由机器学习理论可知,训练一个性能优良的分类器不仅需要一定数量的样本,还需要样本具有代表

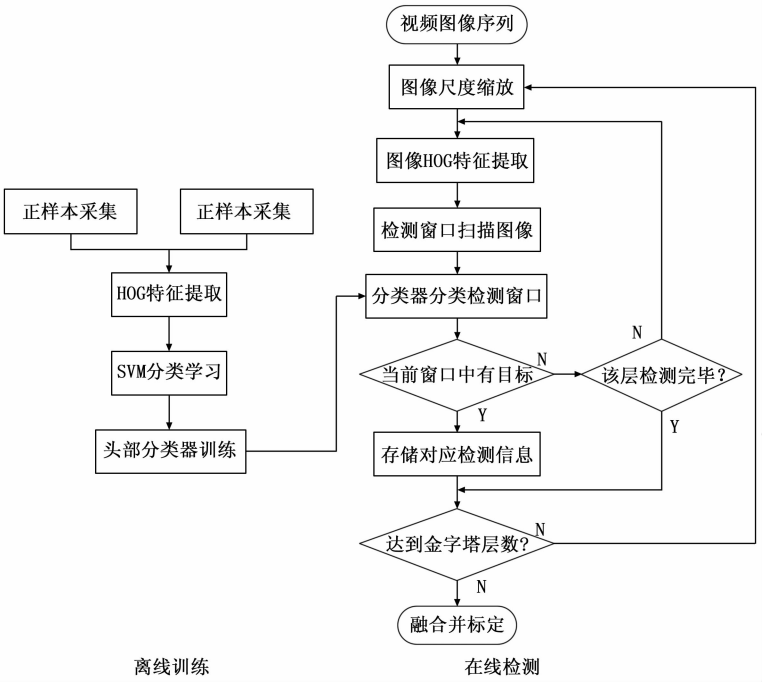


图 1　多尺度协同的人头检测方法流程图

性，即尽可能使其涵盖检测过程中的各种情形。如：应尽量涵盖一定范围内的视野场景、光照、背景等多样性变化条件下的行人头部样本采集情况。摄像头视野下的人体头部虽近似于圆，但头部形状实则因人而异，不同人的头部形状差异明显，在实验条件允许时，应尽量采集不同摄像头下行人的不同发型、不同年龄及性别等特征的人体头部作为正样本，使得到的分类器在离线训练阶段便具有一定的鲁棒性。

2）非头部样本选取实验前，先对背景及周围场景过滤，提取运动目标区域。在作非头部样本选取时，主要集中于人体各部位及附属衣饰，如人的肩部，膝盖、背包等。同时，行人一般活动于较开放的场景，检测中也难免会出现类似行人的运动目标，依现有提取方式，穷举所有可能出现的这类对象较为不易，可尽量准备那些可能出现在视频场景中的类行人头部样本，如花丛、树桩及其它动物等，以完善负样本集。

此外，人工裁剪的样本尺寸大小不一，还需对原始正负样本的尺寸归一调整。本次实验采用图像插值算法[5]，将正负样本统一调整为 64（像素）×64（像素）。经过灰度处理及归一化之后的部分正负样本如图 2 所示。

行人头部可视为刚体，但是不同的个体其形状差异性仍较为明显。不同的人，发型、头饰等也有差异，即使同一人处于摄像头下的位置不同，摄像头读取的头部信息也会有变化。因 HOG 特征在应对图像几何和光学的形变方面能保持很好的不变性，本文以此为特征描述子。

依据支持向量机在做分类决策的优势[6-7]，实验以支持向量机为分类器，并与提取的 HOG 特征结合作分类器的离线训练和实时的头部检测。离线训练阶段，将制作好的正负样本输入到 SVM 模型中，获取该训练后的分类器。实验仅需判断人头及非人头对象，则只需训练一个二分类器；实时检测阶段，将视频序列输入到已训练好的分类器中，通过设定的各参数，使检测窗口在经尺度变换后的不同层作扫描检测，并分类决策。不同的层可能会对同一人头重复检测，一个窗口也可能会因检测到多个人头对象造成的窗口重叠等问题，本文通过融合技术对此优化处理。

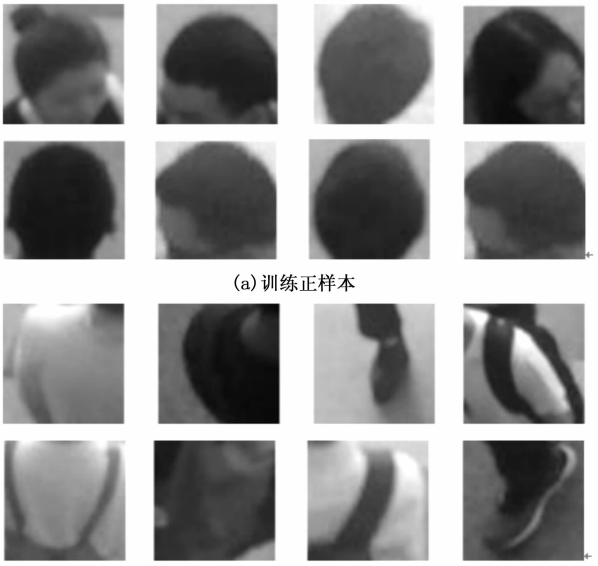
1.2　目标在线检测

1.2.1　多尺度分析

不同视角观测到的自然界的对象呈现出不同形态，机器视觉很难分析出未知场景中物体的尺度[8]，因此，有必要考虑图像同时在多尺度下的分析描述。本文选用高斯金字塔变换处理这一问题。一幅图像的金字塔是一系列以金字塔形状排列的分辨率逐步变换的图像的集合[9]，将图像与一系列大小不同的高斯核作卷积处理，得到图像的多尺度表达，使一幅图像按一定的缩放系数 α 作出变换，实现图像的多尺度变换。此模型的建立可分两步进行：先通过一个低通滤波器的平滑，再对平滑之后的图像作抽样和插值操作，得到按比例缩小或放大的图像，如图 3 所示。序列中的第 i 层图像 L_i（x，y）与相邻的第 $i-1$ 层图像 L_{i-1}（x，y）之间的关系如下：

$$L_i(x,y)=\sum_{p=-2}^{2}\sum_{q=-2}^{2}w(p,q)L_{i-1}(2x+p,2y+q) \quad (1)$$

其中：$L_i(x,y)$ 为第 i 层金字塔图像，$L_0(x,y)$ 为原始输入的视频图像，作为高斯金字塔的第一层，为一个 5×5 的具有低通特性的窗口函数，令 $\overline{h}(p)$ 为高斯密度



(a) 训练正样本

(b) 训练负样本

图 2　部分正负训练样本

分布函数，需满足归一性、对称性、奇偶性等贡献性的约束条件，一个窗口函数 w（p，q）可以表示如下：

$$w(p,q)=\frac{1}{256}\begin{bmatrix}1&4&6&4&1\\4&16&24&16&4\\6&24&36&24&6\\4&16&24&16&4\\1&4&6&4&1\end{bmatrix} \quad (2)$$

由此，可以实现由 L_0，L_1，…，L_N 构成高斯金字塔层级模型。

传统的 HOG 采用单尺度固定大小的检测窗口目标的检测。由于行人通常在运动，很难获取监控视野中不同时刻行人

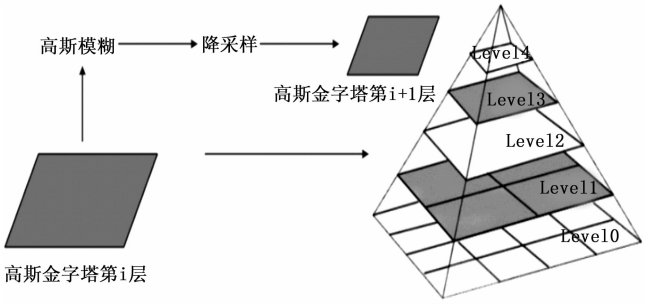


图 3 图像金字塔模型及变换

的大小变化信息。分类器的训练前，已选定了大小合适的样本，而检测窗口的大小取决于训练样本的大小，且样本大小不易改变，这使得检测窗口的大小也相对固定。检测前先对视频序列作高斯金字塔式的变换，这样，原先同一帧中大小不同的人头也可能在某一变换的尺度空间中被检测。本文选定合适的缩放比，对输入的视频序列作高斯金字塔式缩放变换，再对变换后共 5 层不同尺度的图像作扫描检测。

1.2.2 窗口融合分析

窗口融合目的是为得到清晰准确的检测结果，也可减少检测的计算量。人的头部形状的面积趋于相对统一的范围，因此可通过面积阈值过滤检测中可能出现的面积过大或过滤等不符合检测结果的窗口。

实验检测的人头较多，窗口类型较复杂，在 SVM 作分类决策时，人头的形状和大小不一，每帧作高斯金字塔变换后，同一尺度下，当前帧中所有人头被检测的可能性不大。作多尺度检测时，易出现标定多个检测窗口，即窗口重叠的现象，为有效处理较多的检测窗口，且需保证检测结果的准确性，首要原则是降低视频序列中行人头部的漏检率，这可通过调节 SVM 分类器的阈值，以降低检测结果为行人头部的条件来实现；第二，优先处理同层的同类且时空距离较近的窗口；第三，经检测所标定的人头是否准确，及标定的人头位置均以最终检测和融合的结果为准。具体的融合方式为：

1）经每层的检测后，若检测到横坐标相同且重叠的人头，则将其取并集后的头部区域视为一个窗口；

2）经过上述窗口面积阈值处理及同层窗口融合后，取各层人头的选框大小及其质心坐标位置，并映射到在作尺度变换前的对应帧的相应位置。设 P_1，P_2 为任意两个经检测之后的人头标定框，它们是否融合及融合的方式如下：

$$P_{end} = \begin{cases} P_1 \cap P_2, area(P_1,P_2) \geqslant min(P_1,P_2)\times\lambda_2 \\ P_c, min(P_1,P_2)\times\lambda_1 \leqslant area(P_1,P_2) < min(P_1,P_2)\times\lambda_2 \end{cases} \tag{3}$$

若二者间的重叠面积小于它们较小面积的 λ_1 倍，则不作处理，反之，需融合处理。其中，min（P_1，P_2）为取 P_1，P_2 中面积较小者的面积，area（P_1，P_2）表示取检测窗口融合后整体的总面积，P_c 表示经上述处理后得到的新的标定框，其质心坐标可取融合前的两检测框整体的质心，其边长 C 的计算方法为：

$$C = L \times areaRATE' \tag{4}$$

其中：L 是融合前的 P_1，P_2 两个检测框各自边长组成的向量，areaRATE 是这两个检测框各自的面积分别占它们叠加后的总面积的百分比向量；

3）经上述处理后，将融合后得到的最终的检测框的大小、位置等属性映射到在作金字塔变换前的对应帧的相应位置，标定出最终的检测结果。

2 实验及结果分析

当前实验环境为 Win7 32 位系统，Pentium（R）Dual－Core CPU T4300 @ 2.10 GHz 2.10 GHz，Matlab R2014a。正负样本数量分别为 903 和 1098 个。测试视频是某幢大楼一楼大厅某时段的监控信息，其分辨率为 640×480，任截取其中 2880 帧作为实验素材。实验中选取的样本尺寸大小为 64×64 像素，每 8×8 的像素大小组成一个 cell 单元，每 2×2 个 cell 单元组成一个块，一个 block 块的大小为 16×16。而每个 cell 有 9 个特征，则每个块内有 4×9＝36 个特征，实验中设定步幅长度为 8 个像素，则水平和垂直方向都有 7 个扫描窗口，易计算得，64×64 的图片共有 36×7×7＝1 764 个特征。

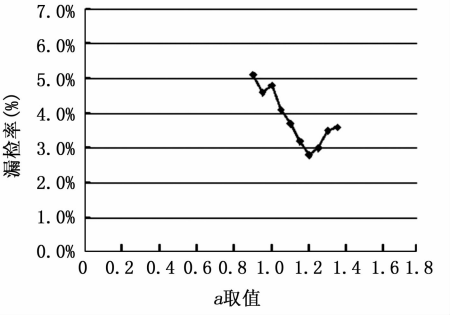


图 4 a 值与漏检率的关系

经多次试验发现，当高斯金字塔缩放系数 $\alpha = 1.2$ 时，漏检率最小，仅为 2.8%，如图 4 所示。窗口融合过程中令阈值 $\lambda_1 = 0.45$，$\lambda_2 = 0.78$ 时，可达最佳实验结果。用同一测试视频对文献［1］、［3］的方法以及本文方法测试对比，部分实验画面如图 5 所示。

为了衡量本文方法的性能，检验方法也采用 Li[10]，Stewart 等[11]使用的方式，需计算的参数指标为：检测率（DetectionRate，DR），召回率（Recall，RC），准确率（Accuracy，AC）等，计算方式如下：

$$DR = \frac{TP}{TP + FP} \tag{5}$$

$$RC = \frac{TP}{TP + FN} \tag{6}$$

$$AC = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

其中：TP（True Positive）表示真阳性，FP（False Positive）表示假阳性，TN（True Negative）表示真阴性，FN（False Negative）表示假阴性。经统计分析，得出三组实验数据如表 1 所示。

表 1 三种检测方法的实验数据（%）

	文献[1]的方法	文献[3]的方法	本文方法
检测率	88.4	86.7	91.2
召回率	89.7	91.3	94.4
准确率	93.4	90.8	96.6

通过上述图 5 的比较分析可知，文献［1］中的方法可检

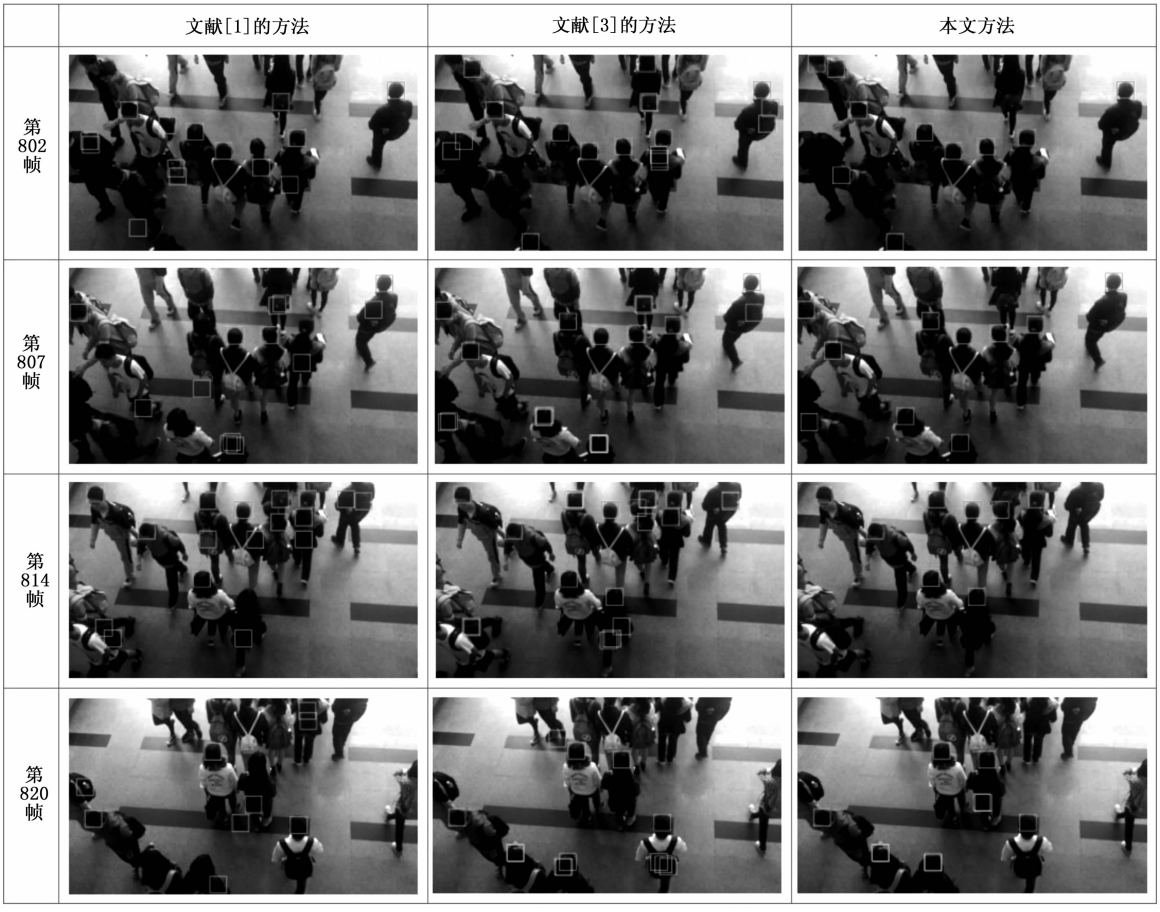


图 5　三组检测方法的部分视频画面

测到实时视频中的目标对象，但漏检较易出现。文献［3］中，当视频画面中出现的人头对象较多时，误判率较高，同一可能的目标对象会被多次检测，使得视频画面中出现多次重叠标定现象，同时也有漏检现象。而本文提出的检测方法在这些方面显然有很大改善，检测结果的标定清晰，在降低漏检率的同时提高了检测的准确性。从表 1 中的数据可知，与前两种检测方法相比，本文方法有效提高了检测率、召回率、准确率等参数指标。实验表明，该方法对于实时视频场景中的人头检测在精确度、检测效率方面有较好改善。

3　结论

多尺度的 HOG 检测继承了 HOG 的优点，高斯金字塔的使用，改善了传统的 HOG 特征的尺度不变性。针对传统的 HOG 特征的单一性问题，本文对输入视频序列先作高斯金字塔式的缩放处理，对缩放后的每层图像作人头的扫描检测。针对传统滑动窗口技术经检测所标定的画面中的重叠现象，本文在有效的检测出实时视频中的人头对象的同时，采用窗口融合方法很好的解决了这一问题，检测效率也有了较大的提升。后续工作将从样本库的完善，检测速度的提升等方面进行。

参考文献：

[1] Zhou T，Yang J，Loza A，et al. Crowd modeling framework using fast head detection and shape－aware matching［J］. Journal of Electronic Imaging. 2015，24（2）：19.

[2] Guan Y，Huang Y. Multi－pose human head detection and tracking boosted by efficient human head validation using ellipse detection［J］. Engineering Applications of Artificial Intelligence. 2015，37：181－193.

[3] Aziz K，Merad D，Iguernaissi R，et al. Head detection based on skeleton graph method for counting people in crowded environments［J］. Journal of Electronic Imaging. 2016，25（1）：13012.

[4] Dalal N，Triggs B. Histograms of oriented gradients for human detection［A］. IEEE Computer Society Conference on Computer Version and Pattern Recognition（CVPR’05）［C］. 2005：886－893.

[5] 郭翰庭. 基于多方向和边缘保持的图像插值算法研究［D］. 成都：西南交通大学，2015.

[6] Meyer D，Wien F T. Support vector machines［Z］. The Interface to libsvm in package e1071. 2015：1－8.

[7] Harris T. Credit scoring using the clustered support vector machine［J］. Expert Systems with Applications. 2015，42（2）：741－750.

[8] 杨　扬. 基于多尺度分析的图像融合算法研究［D］. 北京：中国科学院大学，2013.

[9] Yadav A R，Anand R S，Dewal M L，et al. Gaussian image pyramid based texture features for classification of microscopic images of hardwood species［J］. Optik－International Journal for Light and Electron Optics. 2015，126（24）：5570－5578.

[10] Li B，Zhang J，Zhang Z，et al. A people counting method based on head detection and tracking［A］. Smart Computing（SMARTCOMP）International Conference on［C］. 2014：136－141.

[11] Stewart R，Andriluka M. End－to－end people detection in crowded scenes［J］. Computer Scienct. 2015，1506. 04878：25－26.