

## 综合表观和运动模式的视频轮椅检测方法<sup>①</sup>

李同治<sup>②</sup> 丁晓青<sup>③</sup> 方 驰

(智能技术与系统国家重点实验室 北京 100084)

(清华信息科学与技术国家实验室 北京 100084)

(清华大学电子工程系 北京 100084)

**摘 要** 提出了一种通过综合利用表观和运动模式从视频中检测轮椅(包括坐轮椅的人)的方法。该方法首先使用两种特征描述轮椅的表观特性,并结合利用基于统计学习得到的分类器快速地排除非目标区域。针对由于直立行人和坐在轮椅人之间的表观相似性,轮椅检测易受直立行人干扰的情况,进一步使用运动信息来区分轮椅和正常行人,降低了虚警。在收集的数据上测试了所提方法的性能,实验结果表明提出的方法能够以接近实时的速度进行检测,并且能够有效地降低正常行人的干扰。

**关键词** 轮椅检测,运动模式,级联检测,类 Haar 特征,梯度方向直方图(HOG)特征,光流

## 0 引 言

随着社会物质文明的发展,在一些人流密集场所,如机场、地铁站和大型购物商场等,需要将坐轮椅的残疾人从人群中鉴别出来,为其提供各种服务,使其与健全人平等地享受生活。对残疾人的鉴别,需要用“轮椅检测”技术,这里“轮椅”特指轮椅和乘坐轮椅的残疾人。轮椅检测需要将轮椅从复杂背景中鉴别开,尤其是需要鉴别直立行走或者站立的正常人,然而正常人和残疾人之间的外观差别很小,特别是在正面视角下,坐在轮椅上除使身高变矮外,在表观上与正常人非常接近。因此,轮椅检测的困难是把轮椅从复杂背景和正常行人中鉴别出来。

国内外关于视频中轮椅检测的相关文献较少。文献[1]提出了在标定好摄像机参数的环境中检测轮椅的方法,该方法首先使用背景建模的方法去除背景,然后在前景区域使用肤色检测人脸,通过霍夫变换检测椭圆的方法检测轮子,然后综合两种信息做出最后的判断。然而轮子检测在复杂背景中容易受到背景边缘的干扰,另外在正面视角下,轮子基本上不可见,限制了其应用范围。Huang 等<sup>[2]</sup>使用梯度方向直方图(histograms of oriented gradients,

HOG)特征描述轮椅,使用类似于文献[3]中的级联决策树作为分类器进行判决,且没有利用轮椅独特的运动模式来区分轮椅和正常行人。本文则利用类 Haar 特征和 HOG 特征的优点,且使用运动信息来进一步提高检测性能,提出了一种综合利用表观和运动模式的轮椅检测算法,该方法首先利用轮椅的表观模式逐帧检测,然后对检测到的结果使用运动信息来进一步滤除虚警。基于表观模式目标检测方法已成功运用到人脸、汽车和行人等目标检测中。

## 1 整体流程

引言中已提到,视频中轮椅检测的困难是区分轮椅和复杂背景。由于坐轮椅的人具有和行人相似的表观特性,例如人的头肩、躯干和腿(有时不可见),同时轮椅有其自身的一些表观特点,例如轮子(正面视角下不明显)等,所以可使用这些表观特性将轮椅从普通背景中区分出来。因此,本文首先采用了基于表观模式的检测方法,该方法综合利用了类 Haar 和 HOG 两种不同特征,在保持 HOG 特征检测性能的同时加快了检测速度。轮椅检测的另一个挑战是将轮椅同正常的行人区分开,对于前面提到的几种轮椅的表观特点,正常行人也具有这些特点,

① 973 计划(2007CB311004)资助项目。

② 男,1980 年生,博士生;研究方向:计算机视觉,模式识别;E-mail: azhi207@gmail.com

③ 通讯作者,E-mail: dingxq@mail.tsinghua.edu.cn

(收稿日期:2010-09-20)

两者表观模式具有相似性。因此,仅利用基于表观模式的分类器很难将两者完全区分开。而轮椅与正常行人相比的一个重要区别是其运动模式——各个部位之间具有很高的运动一致性;相反正常行人由于摆臂以及腿的前后交替移动,导致四肢、躯干之间运动方向和幅度一致性差。因此,本文在基于表观模式的检测器后级联了基于运动模式的分类器,用于进一步区分轮椅和行人。

本文采取的检测系统如图 1 所示,由两部分级联构成。第一部分基于表观模式,综合利用类 Haar 和 HOG 特征的优点,采用级联 AdaBoost 分类器,用于从单帧图像中检测轮椅;第二部分从连续的视频中提取运动信息,利用两者在运动模式上的差别,将轮椅同正常行人进行区分。下面对这两个部分分别进行详细的说明。

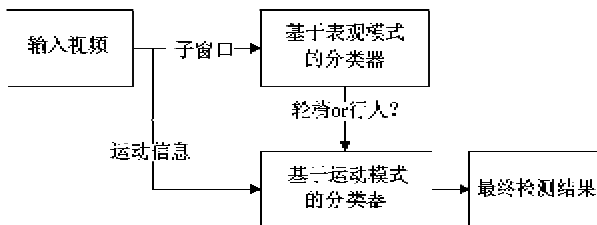


图1 整体处理流程

## 2 基于表观模式的检测方法

基于表观模式的目标检测方法已得到了广泛的研究,其中类 Haar 特征已成功应用于人脸检测<sup>[3]</sup>,其具有运算简单的特点,本文使用它来快速地拒绝掉一些比较简单的背景,然后使用相对复杂的 HOG 特征排除相对复杂的背景区域。HOG 特征在人体检测中取得了优异的性能<sup>[4,5]</sup>,鉴于轮椅与正常人体之间表观模式的相似性,本文应用 HOG 特征来描

述轮椅用于排除相对复杂的背景。基于表观模式的检测器的结构如图 2 所示。

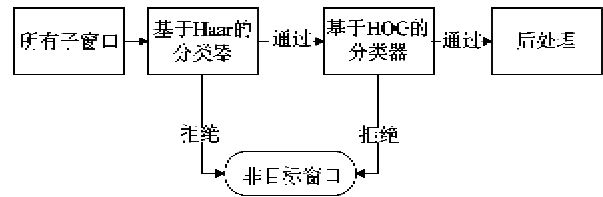


图2 基于表观模式的级联轮椅检测器结构

### 2.1 快速排除背景

Viola 等提出的类 Haar 特征结合级联 AdaBoost 分类器已成功地应用于人脸检测中<sup>[3]</sup>,本文使用和文献[3]一样的检测方法快速排除背景。Viola 等提出的人脸检测算法用两个策略实现背景区域的快速排除:(1)使用积分图表示图像,在计算图像上任意大小的矩形区域内像素的灰度值之和时,只需四次内存读取操作和四次加减法运算,降低了提取类 Haar 特征的运算复杂度;(2)根据目标检测时,非目标窗口/目标窗口数目的不对称性,采用级联分类器结构,级联分类器最初几级使用很少的特征排除掉大量的背景区域,降低了平均运算复杂度。

训练得到的前几级分类器使用到的特征如图 3 所示,从图中可以看出,这里挑选出来的 Haar 特征侧重于比较人体与背景之间的差异,例如双腿与背景之间的差异、脚与地面之间的灰度差异、头与周边背景及肩部的灰度差异等。通过这些能够描述人体与背景之间差异的类 Haar 特征,级联分类器能够快速地排除掉那些灰度单一的背景区域,或者是有灰度变化,但是变化模式与人体形状不匹配的区域,从而达到背景区域快速排除的目标。在训练完使用 Haar 特征的检测器以后,收集通过 Haar 检测器的负样本开始训练基于 HOG 特征的级联分类器。



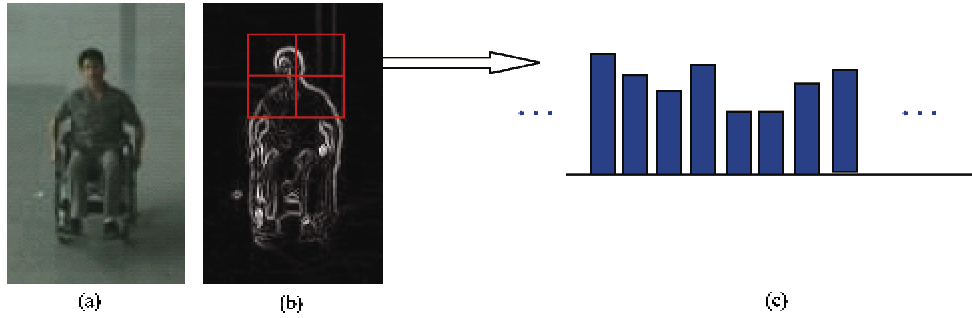
(a)、(b)和(c)依次为 Haar 使用到的前几个特征,(d)和(e)为 HOG 第一、二级使用到的特征

图3 前几级分类器使用得到的特征

### 2.2 使用 HOG 特征的检测

轮廓是人体最显著的特征,而轮廓边缘(包括

位置和方向)的空间分布能够表示人体的形状,例如头肩、上肢的外形以及轮椅的竖直边缘,如图 4(a)



(a)、(b)和(c)依次为原图、梯度图和梯度直方图示意图

图4 HOG特征示意

和图4(b)所示。梯度方向直方图(HOG)特征能够描述局部边缘结构,并且通过位置和方向的量化使其对局部的几何变换(旋转和平移等)具有一定的抵抗能力,另外通过在局部邻域内做归一化处理,使得其对光照变化也具有一定的鲁棒性。文献[4]和[5]已经证明用于行人检测时,HOG的性能要优于类 Haar 特征,由于坐姿的人也具有行人大部分的轮廓特性,HOG 特征也适用于坐姿的人体检测。在使用类 Haar 特征快速地滤除掉背景区域后,本文使用 HOG 特征来排除掉相对更加困难的非轮椅子图像。

特征的另外一个问题是提取特征的位置,不同大小、位置的区域具有不一样的鉴别能力。本文使用类似于文献[3]和[5]方法构造特征池。首先将训练样本归一化到  $W \times H$ , 使用一系列位置  $(x, y)$  和大小  $w \times h$  变化的矩形子窗口  $r_{(x,y,w,h)}$  及对应的 HOG 特征构造特征池,如图3(b)中的矩形框所示。对于每个  $r_{(x,y,w,h)}$  将空间坐标  $(x, y)$  量化到  $2 \times 2$ , 梯度方向  $([0, \pi])$  量化到 9 个量化间隔上,然后利用子窗口内每个像素点梯度幅值的统计梯度方向直方图。这样对于每个子窗口,使用 36 维直方图特征描述。采用类似于文献[3]的方法,使用 AdaBoost 在训练分类器的同时自动挑选具有鉴别信息的特征。

采用积分直方图<sup>[6]</sup>的方法加快 HOG 特征的提取速度。首先,计算图像中每个像素的梯度方向和幅值,并将每个像素的梯度方向按照  $n$  ( $n = 9$ ) 个直方图量化间隔进行量化。然后对每个量化间隔在整幅图像范围内计算幅值积分图。这样,计算任意矩形区域内的梯度直方图时,只需  $4n$  次取内存读取操作和加减运算,加快了特征提取的速度。

分类器采用同文献[3]类似的级联 AdaBoost 结构,其中每级的强分类器定义如下:

$$F(\mathbf{x}) = \sum_{i=1}^T f_i(\mathbf{x}) \quad (1)$$

其中  $f_i(\mathbf{x})$  为定义在每个  $r_{(x,y,w,h)}$  上的分类器,  $\mathbf{x} \in R^d$  为从  $r_{(x,y,w,h)}$  上提取到的 HOG 特征,由于  $d > 1$  (本文中  $d = 36$ ),不能直接使用类似于文献[3]中的单阈值弱分类器,本文首先使用加权的 Fisher 线性鉴别分析(weighted Fisher linear discriminant analysis, WFLD)<sup>[7]</sup>将特征维数降为 1,然后使用基于二次鉴别函数(quadratic discriminant function, QDF)的分类器来表示

$$f_i(\mathbf{x}) = \frac{1}{4} [d_2(\mathbf{x}) - d_1(\mathbf{x})] \quad (2)$$

这里

$$d_i(\mathbf{x}) = \frac{1}{\sigma_i^2} (\mathbf{W}^T \mathbf{x} - \mu_i)^2 + \ln \sigma^2 \quad (3)$$

其中:

$$\mu_i = \sum_{y_j=c_i} w_j (\mathbf{W}^T \mathbf{x}_j) \quad (4)$$

$$\sigma_i^2 = \sum_{y_j=c_i} w_j (\mathbf{W}^T \mathbf{x}_j - \mu_i)^2 \quad (5)$$

这里  $\mathbf{W}$  为投影向量使用 WFLD 学习得到,  $\mu_i, \sigma_i^2$  为第  $i$  类的样本特征投影后的加权均值和方差,  $c_i = \pm 1$  分别代表目标类别和背景类别,  $y_j$  表示样本  $\mathbf{x}_j$  的标签。然后使用 Real AdaBoost<sup>[8]</sup>算法训练分类器的同时挑选具有鉴别性的特征,挑选出来前两个 HOG 特征对应的区域如图3所示。

### 3 基于运动模式的轮椅检测

在引言中已提到,轮椅检测需要将轮椅从普通的背景和行人中鉴别出来。尽管在训练基于 HOG 特征的检测器时,已经将正常直立的行人添加到负样本集中进行鉴别学习,但是还不能完全从图像中排除正常行人的干扰。这一方面是由于行人和坐姿人之间表观相似,特别是正面视角下;另一方面由于行人自身的类内变化大,收集到的行人样本不能覆

盖所有的行人模式,而基于鉴别性学习的分类器依赖于收集到的负样本。图 4 中显示的是分类器使用到的前几个 Haar 特征和 HOG 特征,从这些特征描述的内容可以看出这些特征在正常的行人中也有很强的响应,例如图 3(a) - (c) 中的 Haar 特征,其描述的内容是两腿与背景之间的灰度差,这个在正常直立的行人中间也有很强的响应。经过实验发现仅使用表观模式无法有效地区分轮椅和正常行人。考虑到行人与轮椅在运动模式上有较大的差异,本文采用结合运动信息的方法来区分轮椅和正常的行人。

轮椅和行人之间的运动模式差异主要体现在各个部位之间运动一致性,可以利用这个特点来区分轮椅和正常行人。描述不同部件之间运动一致性的方法有两种,一是检测人体的各个部件,比如上臂、腿和躯干等,然后再跟踪各个部件,使用各个部件的运动轨迹来表示;另外一种是不检测各个部件,直接使用底层的图像块的运动信息来描述。综合考虑最终目的和系统的复杂性,本文采用后一种方式来描述部件之间的运动信息。

本文使用光流来描述物体的运动,光流是底层的运动信息,可以从相邻的两帧图像中直接计算。光流描述了两帧图像局部之间的对应关系,对于运动物体来说,描述的是物体局部区域的运动信息。在通过基于外观的子窗口内计算光流,然后提取特征使用分类器从行人中区分出轮椅。

由于平滑区域计算光流误差大,因此本文使用基于稀疏角点的光流。首先使用 Harris 角点检测算法检测角点<sup>[9]</sup>,然后使用 Lucas-Kanade 金字塔光流计算方法<sup>[10]</sup>进行跟踪。在计算完光流后,统计光流的幅值和方向分布直方图。使用幅度和方向直方图加上幅度的均值和方差来描述物体的运动模式。为了消除物体运动方向对光流方向的影响,首先提取直方图值最大的方向作为主运动方向,对方向直方图做平移处理,使主方向归零。图 5 是行人和轮椅的运动光流特征,从图中可以看出,相比于轮椅,行人具有整体速度快、各个部位之间运动不一致的特点。

由于运动模式只在物体运动时才有效果,而目标没有运动时没有效果,因此在调用分类器之前,首先判断物体的运动幅度,如果运动幅度过小,则不调用该模块;若运动幅度大于某个数值,则调用。

因为学习基于运动模式分类器所需的训练数据来源于视频,需要人工标注大量的视频数据,逐帧标



图 5 行人和轮椅的光流分布示例

注工作量巨大。为此,本文采取了半自动标注的方法标注数据,具体为使用已训练好的基于表观的轮椅和行人检测器自动从视频中标注样本。首先选取只包含轮椅的视频数据,使用基于外观模型的轮椅检测器来标注视频中的轮椅,然后在相邻帧之间提取光流特征,将提取到的特征作为轮椅类别的训练样本;同样,使用训练好的行人检测器采取同样的方法得到行人类别的训练样本。然后使用标注好的这些数据提取运动特征,训练分类器。

## 4 实验及结果

为了实验轮椅检测的性能,我们自己采购轮椅,然后使用摄像机采集视频数据。数据采集的设备和条件如下:共 50 个人,5 把不同的轮椅和 4 个不同的场景,场景采取与地铁站等类似的大厅环境。采集时,每人随机乘坐 5 把轮椅中的两把,然后在不同的场景中行走,使用两个摄像机从不同的角度进行拍摄。最后总共采集到了接近 3 小时的视频。然后按照一定的帧率,从视频中抽取图像进行人工标注,使用标注好的数据训练轮椅检测模型。由于正面轮椅与正常行人最为相似,检测困难,因此本文以正面轮椅检测为例验证提出算法的性能。

### 4.1 仅利用表观模式的检测结果

从采集到的视频数据中,按照 1fps 的帧率提取图像用于标注,使用 2658 (1328 及其左右镜像) 张轮椅图像作为正样本,统一归一化到  $60 \times 80$  的大小,然后以 1700 张不包含轮椅的图像作为负样本。训练分为两个阶段,第一个阶段训练级联 Haar 分类器,本文使用 OpenCV<sup>①</sup> 中的程序进行训练,最后得

① <http://sourceforge.net/projects/opencvlibrary/>

到了一个 18 级的 Haar 级联检测器。然后实验使用不同级数的 Haar 和 HOG 特征联合级联训练。得到的测试结果见表 1。其中 Haar 检测器挑选出的前几个特征如图 3(a)-(c) 所示。HOG 使用到的负样本包括通过 Haar 检测器和一些正常行人数据, HOG 选出来的前几个特征如图 3(d) 和(e) 所示。

从表 1 中可以看出, 使用类 Haar 特征的检测方法整体性能不如使用 HOG 和 Haar 两种特征结合的方法, 但是在 HOG 特征前面使用 Haar 特征后其可以快速地拒绝掉大量非目标窗口, 加快运算速度, 本文使用到的部分测试数据如图 6 所示, 图中的矩形框表示检测到的轮椅位置。

表 1 基于表观模式的测试结果

| 级联结构           | 虚警数目<br>(个) | 检测数目<br>(检测率) | 检测时间<br>(ms) |
|----------------|-------------|---------------|--------------|
| Haar(18)       | 115         | 148(82%)      | 10           |
| Haar(18) + HOG | 27          | 146(81%)      | 49           |
| Haar(1) + HOG  | 15          | 177(98%)      | 82           |
| Haar(2) + HOG  | 13          | 177(98%)      | 66           |

注: Haar 后面() 内的数字表示 Haar 检测器级联的层数



图 6 大厅场景下轮椅检测结果示意图

#### 4.2 基于运动模式的轮椅和行人鉴别

人工挑选了包含 20 段只包含轮椅的视频和 16 段只包含行人的视频作为训练数据, 利用前面提到的半自动标注方法, 基于表观模式的检测器标注数据用于训练分类器。测试数据包含 6 段视频数据, 总长 5.5min, 其中每段视频里面既包含轮椅又包含行人, 然后进行人工标注, 里面总共包含 2516 把轮椅。

提取运动特征时, 首先在检测到的目标区域内使用 Haris 角点检测方法检测角点, 使用这些角点描述检测到的目标, 然后使用分层的 Lucas-Kanade 方法在下一帧中寻找这些角点的对应点, 计算这些角点的运动矢量, 最后统计运动矢量的幅度和方向分布直方图。特征提取参数设置如下: 将光流的方

向 $([-\pi, \pi])$ 均匀量化到 16 个量化间隔上, 将光流幅值量化到 20 个量化间隔上, 然后添加上光流幅值的均值和方差, 总共 38 维的特征向量。

由于轮椅和正常行人之间运动模式差异较大, 本文使用简单的改进的二次鉴别函数(modified quadratic discriminant function, MQDF)<sup>[11]</sup>分类器鉴别行人和轮椅, 最终的测试结果如表 2 所示。

表 2 结合运动模式前后的检测性能对比

| 运动模式 | 漏检率(%) | 虚警数目(个) |
|------|--------|---------|
| 无    | 14.1   | 17      |
| 有    | 15.7   | 2       |

从表 2 中可以看出, 在添加了基于运动模块的行人和轮椅鉴别后, 虚警数目从 17 个(平均 3 个/min)降为 2 个(0.36 个/min), 其中运动的行人被全部滤除掉, 剩下的虚警是站立的行人, 使用运动模式无法排除, 证明了运动模式对于排除掉正常行人干扰的有效性。表 2 中的检测率低于表 1 中的检测率, 这是由两方面的原因造成的: 一方面是由于用于训练检测模型的数据少; 另外一方面是由于视频中轮椅的运动方向的变化, 造成摄像机和轮椅之间相对视角变化, 超出了训练数据覆盖的范围造成了漏检。这个可以通过增加训练数据和加入跟踪模块提高性能。

## 5 结 论

轮椅检测最大的困难是由于复杂背景和正常行人的干扰导致轮椅和非轮椅两类模式之间的差异小。另外, 由于人体衣服多样性的影响, 轮椅(包括坐轮椅的人)本身类内表观模式变化大, 仅利用一种表观特征很难快速地将轮椅同背景进行区分。本文提出的方法综合应用类 Haar 和 HOG 两种特征, 在保持检测速度的同时提高了检测性能。这个方法可以推广到其它的目标检测方法中, 例如行人检测中, 先利用类 Haar 特征快速排除掉平滑的背景区域, 然后使用 HOG 特征排除纹理复杂的背景区域。

由于正常行人和坐在轮椅上的人之间的轮廓相似性, 仅利用表观特征很难有效地将两者区分开; 然而相比于表观上的相似性, 两者在运动模式上有较大的差异。本文提出的基于运动模式的判别方法能够有效地区分正常的行人和乘坐轮椅的人, 该方法还可以应用到其它需要鉴别不同运动模式的人体的

场合,例如智能交通中,骑自行车或者摩托车的人和正常行人的鉴别等。

从实验中可以看出,在使用基于运动模式的分类器后,过滤掉了正常行人的干扰,但是视频中的检测率相对较低,这主要是由于训练数据少和轮椅的姿态变化多引起的,可以通过增加训练数据的方法解决,也可以通过对轮椅整体跟踪,然后融合跟踪和检测信息提高检测性能。另外,本文仅实验了正面轮椅检测,在加入其它视角的轮椅检测后,检测速度明显变慢,如何利用不同视角轮椅之间的共同特点提高检测性能,也是下一步的研究目标。

#### 参考文献

- [ 1 ] Ashish M, Niels D V L, Mubarak S. Wheelchair detection in a calibrated environment. In: Proceedings of the 5th Asian Conference on Computer Vision, Melbourne, Australia, 2002
- [ 2 ] Huang C R, Chung P C, Lin K W, et al. Wheelchair detection using cascaded decision tree. *IEEE Transactions on Information Technology in Biomedicine*, 2010, 14(2): 292-300
- [ 3 ] Viola P, Jones M. Robust real-time face detection. *International Journal of Computer Vision*, 2004, 57(2): 137-154
- [ 4 ] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005. 886-893
- [ 5 ] Zhu Q, Shai A, Mei-Chen Y, et al. Fast human detection using a cascade of histograms of oriented gradients. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, USA, 2006. 1491-1498
- [ 6 ] Porikli F. Integral histograms: a fast way to extract histograms in Cartesian spaces. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005. 829-836
- [ 7 ] Laptev I. Improvements of object detection using boosted histograms. In: Proceedings of British Machine Vision Conference, Edinburgh, Scotland, 2006. 949-958
- [ 8 ] Robert E S, Yoram S. Using confidence-rated predictions. *Machine Learning*, 1999. 37: 297-336
- [ 9 ] Harris C, Stephens M J. A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference, Manchester, UK, 1988. 147-152
- [ 10 ] Bouguet J Y. Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm. [http://robots.stanford.edu/cs223604/algo\\_tracking.pdf](http://robots.stanford.edu/cs223604/algo_tracking.pdf); Intel Corporation Microprocessor Research Labs, 2000
- [ 11 ] Kimura F, Takashina K, Tsuruoka S, et al. Modified quadratic discriminant functions and the application to Chinese character recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, 9(1): 149-153

## A novel wheelchair video detection method based on a combined appearance and motion pattern

Li Tongzhi, Ding Xiaoqing, Fang Chi

(State Key Laboratory of Intelligent Technology and Systems, Beijing 100084)

(Tsinghua National Laboratory for Information Science and Technology, Beijing 100084)

(Department of Electronic Engineering, Tsinghua University, Beijing 100084)

#### Abstract

This paper proposes a novel method to detect wheelchairs in video using the combination of appearance and motion pattern. An appearance based detector is used to detect objects frame by frame, where a cascaded detector is designed using the Haar-like feature and the HOG (histograms of oriented gradients) feature. A motion based classifier is proposed further to distinguish appearance-similar wheelchairs and pedestrians, which usually impedes performance of appearance based detectors. The accuracy and the efficiency of the proposed method were validated by extensive experiments.

**Key words:** wheelchair detection, motion pattern, cascade detection, Haar-like feature, histograms of oriented gradients (HOG), optical flow