

## 基于视频的行人车辆检测与分类

杨 阳<sup>1</sup>, 唐慧明<sup>1,2</sup>

(1. 浙江大学信息与通信工程研究所, 杭州 310027; 2. 浙江省综合信息网技术重点实验室, 杭州 310027)

**摘 要:** 针对传统智能监控中行人车辆检测与分类算法存在目标分割不完整、分类准确率低等问题, 提出一种基于视频的行人车辆检测与分类算法。利用领域信息动态调整置信区间构造混合高斯模型, 采用卡尔曼滤波预测目标下一帧的位置。通过自适应 EM 聚类方法提取目标长宽比和面积作为特征, 将目标分为行人和车辆。在模型估计过程中假设相邻帧目标做匀速直线运动, 推导出目标面积变化满足线性关系, 并对目标跟踪和分类进行修正, 进一步提高检测准确性。实验结果表明, 该算法的人车检测准确率达到 90% 以上, 分类准确率达到 80% 以上。

**关键词:** 行人车辆检测; 智能监控; 运动目标检测; 目标跟踪; 目标分类; 模型估计

**中文引用格式:** 杨 阳, 唐慧明. 基于视频的行人车辆检测与分类[J]. 计算机工程, 2014, 40(11): 135-138.

**英文引用格式:** Yang Yang, Tang Huiming. Pedestrian-vehicle Detection and Classification Based on Video [J]. Computer Engineering, 2014, 40(11): 135-138.

## Pedestrian-vehicle Detection and Classification Based on Video

YANG Yang<sup>1</sup>, TANG Huiming<sup>1,2</sup>

(1. Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China;

2. Zhejiang Provincial Key Laboratory of Information Network Technology, Hangzhou 310027, China)

**[Abstract]** Aiming at the problem of incomplete target segmentation and low classification accuracy of traditional pedestrian-vehicle detection and classification algorithm in intelligent monitoring, this paper presents a pedestrian-vehicle detection and classification algorithm based on video. The algorithm dynamically adjusts confidence intervals for constructing Gaussian mixture model using neighborhood information, and uses the Kalman filter to predict the position of the target in the next frame. It extracts the target aspect ratio and area through adaptive EM clustering as a feature, then divides target into pedestrians and vehicles. Assume that target makes the uniform linear motion in adjacent frame and derive the target area to meet the linear relationship change. Thus target tracking and classification can be modified to improve the detection accuracy in the end. Experimental result show that the algorithm detection rate is over 90% and classification rate is over 80%.

**[Key words]** pedestrian-vehicle detection; intelligent surveillance; motion object detection; object tracking; object classification; model estimation

**DOI:** 10.3969/j.issn.1000-3428.2014.11.027

### 1 概述

智能监控系统<sup>[1]</sup>可不在人为干预情况下对摄像机记录的视频序列进行分析, 并且对其中的行人和车辆等目标进行检测和分类, 从而实现视频信息检索。目前常用的行人车辆检测分类方法可以分为: 基于监督学习的方法和基于运动分割的方法。

基于监督学习的方法需要预先对分类器进行训练, 例如, 文献[2]提取行人的 HoG 特征放入 SVM

分类器进行训练, 类似的有 Harr 特征与 Adaboost<sup>[3]</sup>算法, 这类方法的优点在于准确度高、摄像机晃动影响小, 但是缺点在于时间复杂度大, 而且需要事先进行训练, 场景变换后需要重新训练。

基于运动分割的方法要求摄像机固定, 对运动的物体进行检测分类。例如, 文献[4]通过帧差法检测运动物体, 提取运动物体长宽比作为特征进行分类, 然后利用跟踪将目标在每一帧获得的分类结果进行统计, 最终输出统计结果的总数。然而, 由于采

**基金项目:** 国家科技重大专项基金资助项目(2010ZX03004-003-01); 中央高校基本科研业务费专项基金资助项目(2012FZA5008)。

**作者简介:** 杨 阳(1988 -), 男, 硕士研究生, 主研方向: 视频图像处理, 智能监控; 唐慧明, 副教授。

**收稿日期:** 2013-12-18    **修回日期:** 2014-01-10    **E-mail:** thm@zju.edu.cn

用固定阈值进行目标分类,缺乏自主学习能力,因此对应用场景有严格限制。文献[5]对上述方法进行了改进,对运动目标的特征进行聚类,形成树分类器,能够得到较好的结果,然而该算法也存在过训练、收敛速度慢的问题。

本文针对静止场景的视频序列,提出一种基于运动分割的视频内容检索算法,以提高检测和分类的准确性。

## 2 行人车辆检测与分类原理

本文算法主要包括 4 个部分:运动目标检测,目标跟踪,目标分类和模型估计,如图 1 所示。

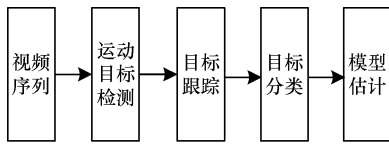


图 1 行人车辆检测与分类算法流程

### 2.1 运动目标检测

运动目标检测作为视频分析的最初阶段,其结果好坏直接影响后期处理的准确性。一个好的检测算法需要能够处理光线亮度的变化、动态背景、前景目标与背景相近等问题。文献[6]提出混合高斯模型来解决动态背景的问题,但是像素点之间是独立关系,没有利用到邻域信息。文献[7]采用无参数窗函数估计,得到的概率模型更贴近实际,但时间复杂度高。文献[8]采用单高斯模型与马尔科夫随机场(Markov Random Fields, MRF)融合对运动目标检测进行改进,能够得到更加完整的前景,但是仍然存在噪声抑制差的缺点<sup>[9]</sup>。

本文利用邻域信息构建一个滤波函数,得到一种混合高斯模型(Gaussian Mixture Model, GMM)的改进算法。

假设像素点  $i$  当前的灰度值为  $x^{(i)}$ ,  $x^{(i)}$  隐含的映射关系为  $f^{(i)}$  (前景和若干个背景)。假定  $f^{(i)}$  条件下,灰度值  $x^{(i)}$  满足高斯分布,即:

$$(x^{(i)} | f^{(i)} = j) \sim N(\mu_j, \sum_j) \quad (1)$$

其中,  $\mu_j$  为高斯分布  $j$  的均值;  $\sum_j$  为  $j$  的方差。

定义像素点  $i$  所在邻域  $N_i$  范围内的滤波函数为:

$$p(f^{(i)}) = \frac{1}{Z} \cdot E(f^{(i)}) \quad (2)$$

其中,  $Z$  为归一化常数;  $E(f^{(i)})$  为势能函数,定义如下:

$$E(f^{(i)}) = \sum_{c \in C} V_c(f^{(i)}) \quad (3)$$

$$V_c(f^{(i)}) = \begin{cases} 0 & f^{(i)} \neq f^{N_i} \\ \beta & f^{(i)} = f^{N_i} \end{cases} \quad (4)$$

其中,  $\beta$  为相互作用系数;  $c$  表示领域  $N_i$  的子群;  $C$  表示子群集合;  $f^{N_i}$  表示像素点  $i$  周围点的映射关系。式(3)的作用在于计算周围像素点的映射与  $i$  点映射的相似性。

在传统混合高斯模型中,运动检测过程等同于求取后验概率  $p(x_i/f^{(i)})$  最大值的问题。对后验进行滤波:

$$p(f^{(i)} | x_i) \cdot p(f^{(i)}) \quad (5)$$

将式(1)、式(2)代入式(5)中得到:

$$p(f^{(i)} = j | x_i) \cdot p(f^{(i)} = j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \cdot \exp\left(-\frac{(x - \mu_j)^2}{2\sigma_j^2}\right) \cdot \frac{1}{Z} \left(\sum_{k=1}^{N_i} \beta_{k,j}\right) \quad (6)$$

如果直接计算式(6)的最大值,则计算过于繁琐,实际中采用判断是否落入置信区间进行近似,得到:

$$|x_i - \mu_j| < N \cdot \sigma_j \cdot \frac{1}{Z} \cdot \sum_{k=1}^{N_i} \beta_{k,j} \quad (7)$$

其中,  $N$  为置信度,取值 2~3。已知传统混合高斯模型的匹配公式为:

$$|x - \mu_i| < N \cdot \sigma_i \quad (8)$$

对比式(7)、式(8)可以看出,改进算法在传统混合高斯模型基础上对置信度进行了修正,利用邻域信息动态地增加或减少匹配该模型的概率。

### 2.2 目标跟踪

为得到更准确的行人和车辆分类结果,需要跟踪目标从出现到消失的所有状态,从而联合各帧信息提高分类正确率。常见的目标跟踪算法有:帧间匹配<sup>[10]</sup>,卡尔曼滤波<sup>[11]</sup>,粒子滤波, Camshift 以及 Estembling Tracking<sup>[12]</sup>等。

本文采用卡尔曼滤波预测目标下一帧位置,同时提取目标的颜色直方图作为帧间匹配特征<sup>[13]</sup>。因为在运动目标检测阶段可以得到每一个目标的掩膜,所以跟踪时不需要在原始图像中进行搜索,只需在前景图中进行目标匹配,这样做可以大大节省跟踪时间。然而,如果目标之间出现遮挡情况,则运动目标检测会将其判断为一个目标,导致上述方法失效。如果目标之间存在遮挡的情况,本文采用 Camshift 算法。

### 2.3 目标分类

目标分类旨在自动判定目标所属类别,实现过程包含 2 个步骤:特征提取和分类判断<sup>[14]</sup>。本文选取目标长宽比和面积作为特征,将目标分为行人和车辆。采用一种自适应 EM<sup>[15]</sup> 聚类方法,实现步骤如下:

(1)将图像划分为若干个 1 616 的区域,每一个区域分别记录覆盖该区域的行人和车辆的信息,包

括面积均值、面积方差以及样本数量,并假设该区域内的行人和车辆目标面积满足高斯分布。

(2)EM 初始化阶段:采用一个固定长宽比阈值来判定行人和车辆,例如  $\text{height}/\text{width} > 1$  表示行人,否则为车辆,相应地记录行人和车辆的面积平均值。

(3)EM 估计阶段:对于每一个区域,当有新目标进入时,计算目标面积值所匹配的高斯分布,从而得出分类结果,相应地更新该类的参数。

通过上述步骤,随着样本数量的增加,高斯分布的参数逐渐收敛<sup>[11]</sup>,分类效果逐渐改善。

2.4 模型估计

模型估计是对历史信息的一种统计应用,要求假设能够逼近实际情况。在相邻帧,由于运动时间很短、位移较小,本文假设目标做匀速直线运动,可以证明其面积变化呈线性关系。

设  $(X, Y, Z)$  和  $(X', Y', Z')$  分别表示第  $k$  帧和第  $k + 1$  帧时刻目标的世界坐标位置。目标的平移、旋转运动可以用式(9)表示:

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \quad (9)$$

忽略目标的深度信息,假设目标为一平面,即:  
 $aX + bY + cZ = 1$  (10)

如果运动目标与摄像机的距离远大于目标本身的深度,即为正投影模型,则存在如下近似关系:

$$(x, y) = (X, Y), (x', y') = (X', Y') \quad (11)$$

将式(11)代入式(9)、式(10)得到:

$$x' = a_1x + a_2y + a_3, y' = a_4x + a_5y + a_6 \quad (12)$$

分别取第  $k$  帧和第  $k + 1$  帧时目标矩形框的左上角上下顶点:

$$\begin{cases} p_1(x_1, y_1), p_2(x_2, y_2) \\ p'_1(x'_1, y'_1), p'_2(x'_2, y'_2) \end{cases} \quad (13)$$

因为存在  $x_1 = x_2, x'_1 = x'_2$  的关系,将其代入式(12)和式(13)中得到,矩形高度满足  $h' = a_2 \cdot h$ 。同理,矩形宽度满足  $w' = a_1 \cdot w$ 。所以,矩形框面积为  $S' = kS$ ,呈线性关系。

利用上述结论,可以对目标分类和目标跟踪进行修正。目标分类阶段图像特定区域存在特征样本少或者收敛速度慢的问题,由于相邻区域面积满足比例变化,因此可以利用周围块的平均值代替当前块。目标跟踪阶段,由于目标面积变化满足线性条件,因此可以增加目标面积作为卡尔曼滤波器的状态量,从而改善跟踪的准确性。

3 实验结果与分析

实验在 PETS2006 视频库以及真实采集的监控视频上进行,以下介绍各阶段的实验结果。

3.1 运动目标检测实验结果

图 2 展示了本文方法与文献[5-7]方法的实验对比结果。

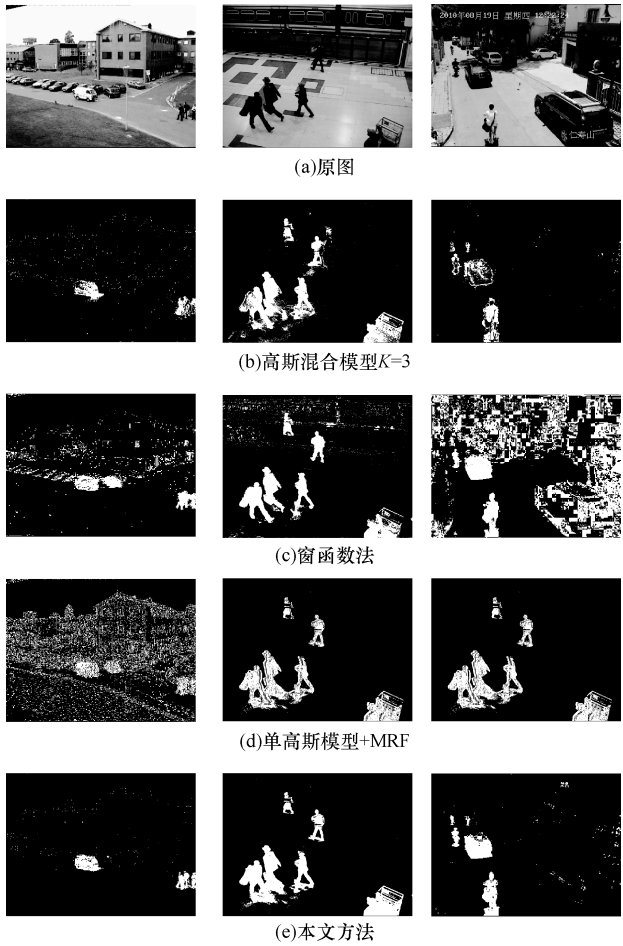


图 2 运动目标检测结果

由图 2 可知,与混合高斯模型相比,本文方法运动目标分割更加完整、噪声影响更小。对比其他常用的运动检测方法,针对不同的室内室外场景,本文方法都能得到较好效果。

表 1 统计了图 2 中 3 个不同场景下,高斯混合模型和本文方法的平均错误率,可以看出虚警率和漏警率都有明显的下降。其中,虚警率和漏警率的定义如下:

虚警率 = 背景点被误检为前景的数量/检测为前景的总数

漏警率 = 被遗漏的前景点数量/真实前景点的总数

表 1 运动检测错误率			%
检测算法	虚警率	漏警率	
混合高斯模型	5.63	6.58	
本文算法	1.88	4.12	



### 3.2 目标分类实验结果

图3展示了目标分类的一帧结果,即使存在遮挡情况,仍然能够得到准确的分类。表2为视频中车辆从出现到消失过程中,每一帧分类结果的统计图,可以看出分类准确度在80%以上。



图3 人车分类结果

表2 人车分类结果的统计

类别	出现次数
车辆	228
行人	19

### 3.3 模型估计实验结果

图4展示了一个做匀速直线运动的行人,面积随时间的变化关系。可以看出面积值可以近似拟合为线性,从而证明本文假设贴近实际情况。

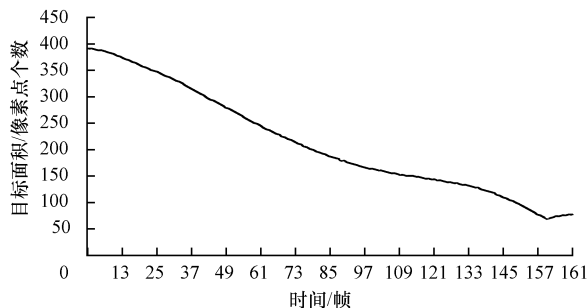


图4 目标面积变化

## 4 结束语

本文提出一种行人车辆检测算法,通过运动目标检测、目标跟踪、目标分类和模型估计等步骤实现。利用该算法能够准确地检索出监控视频中特定目标的信息。与传统算法相比,优势在于能够大幅提高准确度,并且由于采用自适应学习与分类方法,应用场景更加广泛。在今后工作中,将通过统计目标在图像上各区域的特征信息对场景进行建模,估计图像深度,自动学习场景参数。

### 参考文献

- [1] 刘治红,骆云志. 智能视频监控技术及其在安防领域的应用[J]. 兵工自动化,2009,28(4):75-78.
- [2] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2005:25-28.
- [3] Viola P, Jones M, Snow D. Detecting Pedestrians Using Patterns of Motion and Appearance [J]. International Journal of Computer Vision, 2005, 63(2):153-161.
- [4] Lipton A J. Moving Target Classification and Tracking from Real-time Video [C]//Proceedings of the 4th IEEE Workshop on Applications of Computer Vision. Princeton, USA: IEEE Press, 1998:8-14.
- [5] Stauffer C. Learning Patterns of Activity Using Real-time Tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8):747-757.
- [6] Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-time Tracking [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Computer Society, 1999:2246-2257.
- [7] Elgammal A, Harwood D R. Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance [J]. Proceedings of the IEEE, 2002, 90(7):1151-1163.
- [8] Kertesz C. Texture-based Foreground Detection [J]. International Journal of Signal Processing of Image Processing and Pattern Recognition, 2011, 4(4):51-61.
- [9] Geman S, Geman D. Stochastic Relaxation Gibbs Distributions and the Bayesian Restoration of Images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984, 6(6):721-741.
- [10] Surendra G, Osama M. Detection and Classification of Vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2002, 3(1):37-47.
- [11] Kalman R E. A New Approach to Linear Filtering and Prediction Problems [J]. Journal of Basic Engineering, 1960, 82(10):35-45.
- [12] Avidan S. Ensemble Tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2):261-267.
- [13] 刘亚,艾海舟,徐光佑. 一种基于背景模型的运动目标检测与跟踪算法 [J]. 信息与控制, 2002, 31(4):315-319.
- [14] 万纓,韩毅,卢汉清. 运动目标检测算法的探讨 [J]. 计算机仿真, 2006, 23(10):221-226.
- [15] Dempster A P, Laird N M, Rubin D B. Maximum Likelihood from Incomplete Data via the EM Algorithm [J]. Journal of the Royal Statistical Society, 1977, 139(1):1-38.

编辑 陆燕菲