

基于改进 ViBe 的多行人检测方法

朴春赫^{1,2}, 潘怡霖¹, 赵海¹, 朱宏博¹

(1. 东北大学 计算机科学与工程学院, 辽宁 沈阳 110819; 2. 金策工业综合大学 电子工程系, 朝鲜 平壤 999093)

摘 要: 利用改进的 ViBe 算法及模板匹配方法对多行人场景进行了目标检测. 为了提高运动目标检测性能, 提出将原 ViBe 算法与帧间差分相结合的方法, 该方法使二值图像中的鬼影比原 ViBe 算法消除更为迅速. 根据待处理像素点周围已完成的前背景分离结果, 利用动态计算模型参数估计前景, 以提高前景目标提取的准确度. 基于 HOG 算法识别运动目标中的行人目标. 实验结果表明, 该方法在没有降低运动物体检测实时性的同时, 明显提高了算法检测的准确度.

关键词: 运动目标检测; 鬼影消除; ViBe; 支持向量机; 行人检测

中图分类号: TP 317.4 文献标志码: A 文章编号: 1005-3026(2016)04-0481-05

Multi-pedestrian Detection Approach Based on Improved ViBe Algorithm

PAK Chun-hyok^{1,2}, PAN Yi-lin¹, ZHAO Hai¹, ZHU Hong-bo¹

(1. School of computer Science & Engineering, Northeastern University, Shenyang 110819, China; 2. School of Electronic Engineering, Kim Chaek University of Technology, Pyongyang 999093, DPRK. Corresponding author: PAK Chun-hyok, E-mail: vlsi126@126.com)

Abstract: The moving object detection of multi-pedestrian scenario was conducted by using the improved ViBe algorithm and the template matching method. The original ViBe algorithm was combined with the frame difference method to improve the performance of moving object detection, which can more rapidly eliminate the ghosts in the result binary image than the original ViBe algorithm. According to the treated foreground result around the treating pixel, the dynamic model parameters were used to estimate the foreground for improving the accuracy of the foreground object detection. The HOG algorithm was used to detect the pedestrians in the moving objects. The results showed that the method does not reduce the real-time performance of moving object detection, and significantly improve the accuracy of the detection algorithm.

Key words: moving object detection; ghost elimination; ViBe; SVM; pedestrian detection

运动目标检测是从视频图像序列中提取出感兴趣的运动目标, 其算法可分为 6 大类, 分别是光流法^[1]、帧间差分法^[2]、背景差分法^[3-5]、聚类分析法^[6]、图像分割法^[7]和点检测法^[8-9]. 由于其他算法对于摄像头位置的变化、光照的变化或背景中存在小幅度震动物体等问题的处理效果不是很理想, 而背景差分法在上述问题中表现效果良好, 因此被广泛认为是最为有效的方法. 背景差分法又分为多帧平均、混合高斯背景建模、像素值聚类及 ViBe (visual background extractor) 算法^[5].

ViBe 算法由于其算法简单、速度快等优点受到广泛重视. 但有时会因引入鬼影等问题降低检测精度, 针对这一问题本文提出了改进的 ViBe 算法.

在行人检测中, Dalal 等^[10]基于 SIFT 尺寸不变特性的启发提出梯度直方图 (histogram of oriented gradients, HOG), 实现目标特征性状的描述, 通过支持向量机 (support vector machine, SVM) 分类器实现目标检测. HOG 描述图像的局部边缘梯度信息, 对小量的偏移和光照变化具有很好的鲁棒性. 缺点是维度较高, 以致其提取速度

收稿日期: 2015-04-13

基金项目: 中央高校基本科研业务费优秀科技人才培养资金资助项目 (N140405004); 辽宁省科技厅重点实验室建设基金资助项目 (LZ201015).

作者简介: 朴春赫 (1976-), 男, 朝鲜平壤人, 东北大学博士研究生; 赵海 (1959-), 男, 辽宁沈阳人, 东北大学教授, 博士生导师.

很慢.

目前随着摄像头技术的快速发展,CCTV 摄像头的图像分辨率比以往图像分辨率有了很大提高,所以对提取的图像直接使用 HOG 会导致系统实时性低下.因此文献[11]先通过运动目标检测技术找出 ROI(region of interesting),再对 ROI 使用 HOG 算法进行行人检测.本文的大体思路虽与文献[11]相似,却通过使用改进的 ViBe 算法及算法适用性技巧,使得检验结果的实时性和准确率明显提高.

1 相关工作

1.1 ViBe 算法

1) 背景建模:背景模型中的每一个像素由 N 个背景样本组成. ViBe 算法使用第一帧图像初始化背景模型.在欧几里得颜色空间中图像 x 位置的像素值为 $V(x)$, V_i 表示索引为 i 的背景样本值^[5].背景模型 M 定义为

$$M(x) = \{V_1, V_2, \dots, V_N\}. \tag{1}$$

2) 前景检测:前景检测通过当前像素与背景模型中对应的历史像素值进行比较,确定当前像素是背景像素还是前景像素.用 $S_R(V(x))$ 表示以 $V(x)$ 为中心, R 为半径的圆形区域, $\# \{S_R(V(x)) \cap \{V_1, V_2, \dots, V_N\}\}$ 表示 $M(x)$ 中的点,处于 $S_R(V(x))$ 区域中的个数,如图 1 所示,其中 C_1, C_2 为二维颜色空间 (C_1, C_2) 的分量.

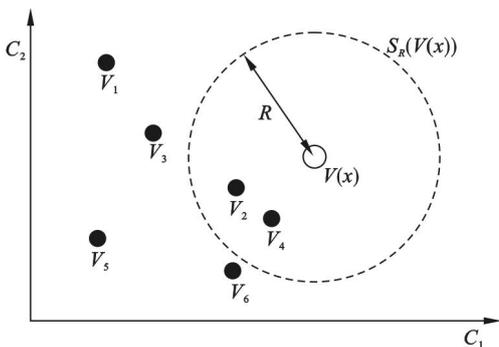


图 1 二维欧式空间中的像素分类

Fig. 1 Pixel classification in 2D Euclidean space

将阈值用 $\#min$ 表示,将 $\#min$ 与 $\# \{S_R(V(x)) \cap \{V_1, V_2, \dots, V_N\}\}$ 进行比较,判断当前像素属于前景还是背景.

3) 模型更新:ViBe 更新采用随机策略.如果像素 $P(x)$ 是背景像素,则随机从 $M(x)$ 中选取一个值用 $P(x)$ 的像素值 $V(x)$ 代替;与此同时,为了使像素邻域空间的一致性得以保持,ViBe 算法采用同样的方法对邻域 $NG(x)$ 中像素的背景模型

进行更新.

1.2 HOG 算法

在基于统计分类的行人检测方法中,最具代表性的是文献[10]提出的基于 HOG 和 SVM 的行人检测方法. HOG 方法可有效地进行行人检测,基于 HOG 的行人检测算法获取 HOG 描述子 descriptor 的方法主要分为 5 个步骤:伽马(Gamma)校正;计算像素梯度;统计单元(Cell)内方向梯度信息;归一化块(Block);生成 HOG 特征向量.对以上步骤进行基于 SVM 的分类器训练,可以训练出指定类型的行人 SVM 分类器,利用该分类器对输入图像进行搜索分类即可实现行人检测.本次实验中使用的 SVM 分类器是 opencv 中已经训练好的行人检测分类器.

2 算 法

2.1 ViBe 算法的改进

1) 鬼影消除:鬼影是一组在运动过程中被检测到的连接点,它不对应于任何真实存在的物体.在 ViBe 算法中,鬼影是由于背景建模过程中将样本中运动目标(即前景)错误地初始化为背景而产生的影像.虽然随着时间推移,鬼影可以通过更新背景模型逐渐消除,但是这一过程相对缓慢.鬼影的存在会降低运动目标检测精度.为实现对鬼影的快速消除,提出下述方法:设 t 时,对已判定为前景像素的点 p 进行帧间差分处理,并设置一个计数器 $TOS(p)$,初始化为 0,如果该点的像素值 $V^t(p)$ 与之前帧对应的像素值 $V^{t-1}(p)$ 的差分结果在预设阈值 T_d 之内,则 $TOS(p)$ 自动加 1,否则, $TOS(p)$ 清零.该过程可以表示为

$$TOS(p) = \begin{cases} TOS(p) + 1, & |d(p)| \leq T_d; \\ 0, & \text{其他.} \end{cases} \tag{2}$$

式中 $d(p)$ 为帧上对应像素之间的像素值差分.

2) ViBe 自适应参数:通过以往的实验发现^[5],当颜色扭曲尺度 $R = 20$ 、样本匹配数 $N_{min} = 2$ 时,在很多情况下都可以获得较好的实验结果.其中,颜色扭曲尺度是用来调整与当前帧中对应点的相似背景模型中样本的变化范围,因此, R 越大背景像素点越多.相反, R 越小则产生变化很小的点也会被分类为前景像素点.介于该思想, Droogenbroeck 等于 2012 年提出了一种自适应参数方法^[12],通过计算每个像素点 p_i 的背景模型的标准差 σ_i ,动态设置模型参数 R ,使其在 $[20, 40]$ 的范围内.但是这种方法的计算量太大,导致检测速度很慢.这种自适应参数方法在实验环境

中实际执行速率照原算法降低了约 1/4,但其方法中相邻像素点具有相似的时空分布的思想却有着较高的参考价值,基于此本文提出了一种改进自适应参数的方法.

运动目标检测算法的目标是能够在正确分割出前景图像的同时可以剔除与实际运动物体无关的前景像素点. 本文考虑了算法的速度和准确度,对原有方法进行了如下改进.

已知处理过的像素的状态会影响正在处理的像素,对于闪烁点也是如此. 为了实现这一点,引入新参数作为周围像素点对当前像素点分类的影响因子,影响因子的设置如图 2 所示.

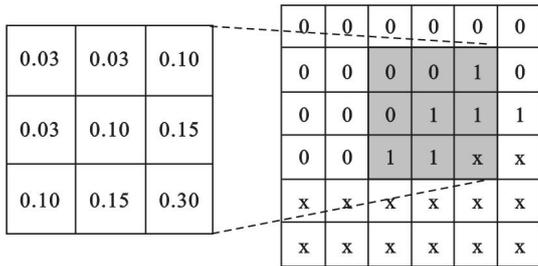


图 2 加权窗口的适配参数

Fig. 2 Weighted window for adapting parameter

设置各个加权窗口的原因:需要保证加权窗口的和约等于 1,同时,假设距离处理点越近,则该点的影响力越大. 鉴于处理速度,处理过的像素点的影响范围设计为 3×3 的矩阵. 通过实验验证,当各个权值按如图 2 所示设定时,其效果最为理想. 正在处理的像素为 3×3 方格中的右下角的像素点. 例如:加权窗口与对应的正在处理点的周围处理过的像素点的二进制状态的卷积为 1,则 $R = 2$; 如果卷积的值为 0.3,则 $R = 20$. R 值范围设定为 $[2, 20]$.

动态半径 R 的自适应系数 T 的具体计算式为

$$T = W \otimes B . \tag{3}$$

式中: W 为加权窗口; B 是相应的处理像素(对于运动目标的准确检验,正在处理像素点的状态假设为 1); \otimes 为卷积运算. 当前像素点 R 和 T 之间的线性关系借鉴了文献[12]中的方法:

$$R = -25.7 \cdot T + 27.7 . \tag{4}$$

2.2 本文提出的算法

ViBe 对前景图像的提取效果及鲁棒性好,但速度较慢,针对这一特点,提出了对输入图像进行压缩预处理. 然后对压缩处理过的图像用改进 ViBe 进行运动目标检测,再对获取的运动目标进行基于 HOG 的 SVM 行人检测. 具体流程图如 3

所示.

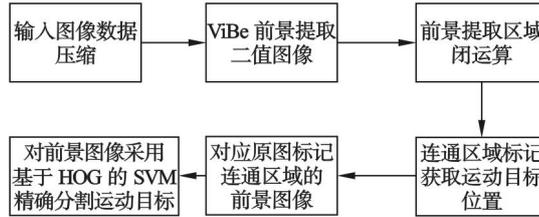


图 3 本文提出的算法过程

Fig. 3 Block diagram of proposed algorithm

具体的算法如下:

- 步骤 1 输入视频序列 $S = \{s_1, s_2, \dots, s_i, \dots, s_n\}$;
- 步骤 2 输入当前视频帧 s_i 并进行压缩;
- 步骤 3 采用改进 ViBe 算法对当前帧进行运动目标区域提取;
- 步骤 4 闭运算去除噪点,填补区域空白;
- 步骤 5 获取运动目标位置,放大运动目标区域进行连通区域标记;
- 步骤 6 对该区域采集 HOG 特征,对原图的 ROI 进行行人检测;
- 步骤 7 重复步骤 2 ~ 步骤 6,直至所有的 ROI 区域均被找出,终止循环并返回计算结果.

图 4、图 5 分别为二值图像的闭运算前后处理对比图和连通域标记法对目标之间独立时分割结果图.

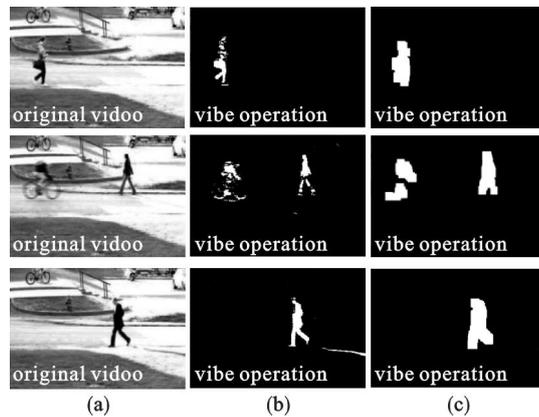


图 4 二值图像的闭运算前后处理对比图

Fig. 4 Foreground images before and after closing operation

(a)—原图; (b)—闭运算前; (c)—闭运算后.

3 实验及分析

3.1 改进 ViBe 算法的运动目标检测准确度

对前景目标的检测是对图像像素二值划分的过程. 二值划分结果的正确性用 3 个指标来评

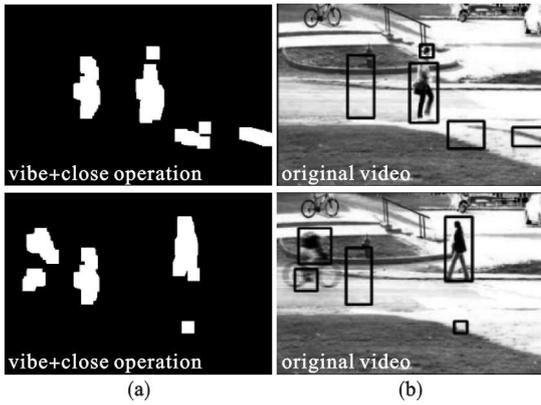


图 5 当目标之间独立时连通域标记法分割结果图
 Fig. 5 Result image processed by component-labeling to independence target segmentation image
 (a) — 压缩图的二值图像; (b) — 原图的标记图.

价^[5], 分别是召回率 (recall)、精确率 (precision) 和它们的调和平均值 F . 3 个评价指标的定义为

$$\text{recall} = \frac{N_{\text{true_result}}}{N_{\text{GT}}}, \quad (5)$$

$$\text{precision} = \frac{N_{\text{true_result}}}{N_{\text{result}}}, \quad (6)$$

$$F = 2 \cdot \frac{\text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}. \quad (7)$$

式中: $N_{\text{true_result}}$ 为检测结果中正确前景像素的个数; N_{result} 为检测结果中前景像素的个数; N_{GT} 为标定结果中前景像素的个数.

实验输入一组连续的图像序列. 对于不同帧

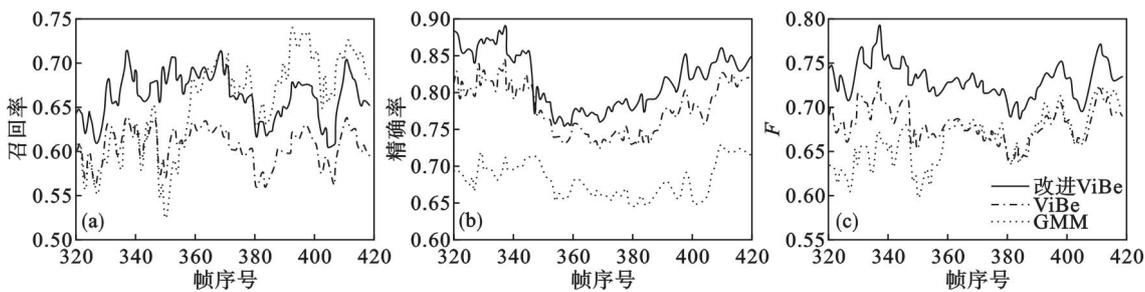


图 7 运动目标检测结果
 Fig. 7 Detection results of moving objects

3.2 改进 ViBe 算法的行人检测实效性分析

为了验证所提出算法的实时性, 使用了 3 种方法进行行人检测效果对比. 这 3 种方法分别是使用 HOG 算法对全图像进行行人检测, 采用 ViBe 和 HOG 相结合算法对无压缩图像中的行人目标进行检测, 以及本文方法. 实验平台为 CPU Core I5 4570 3.2 GHz, DRAM 4G. 实验中利用的行人模板为 INRIA Person Database^[10]. 图 8 为在

大小的多环境下连续图像进行了前景分割, 对比 ViBe 与改进 ViBe 的提取效果. 对于 ViBe 算法, 利用文献[5]中的参数值, 即背景模型的大小 $N = 20$, 颜色扭曲尺度 $R = 20$, 样本匹配数 $N_{\text{min}} = 2$, 二次抽样参数 $\varphi = 16$.

图 6 为 2 个算法分辨率为 360×240 的连续图像序列时的鬼影消除效果. 图 7 为 3 个算法 (ViBe、改进 ViBe 和 GMM) 适应于连续图像序列时, 通过式(5)~式(7)估计的结果. 由图 7 可知, 改进的 ViBe 在与基础 ViBe 算法和 GMM 算法相比, 在精确率和 F 值等评价体系中拥有着较大的优势, 特别是综合估计参数 F 值要明显优于其他方法.

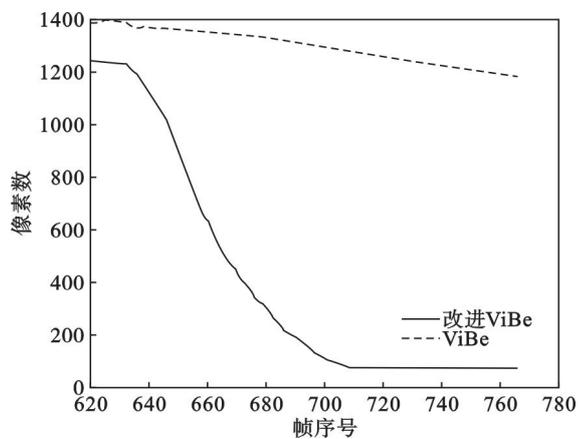


图 6 ViBe 改进前后的鬼影像素数
 Fig. 6 Number of ghost pixels before and after improved ViBe

分辨率为 768×576 的连续图像序列中得到的行人检测结果.

图 9 为在分辨率为 768×576 的连续图像序列中所得到的实时性对比结果. 结合图 8 可知, 本文算法的实时性受到了场景复杂度的制约, 在场景中存在的行人运动目标个数与算法速度成反比, 但是相较于其他方法本文方法在检测实时性方面存在较为明显的优势.



图8 行人检测结果

Fig. 8 Detection results of pedestrian

(a)—HOG; (b)—HOG + ViBe; (c)—本文方法.

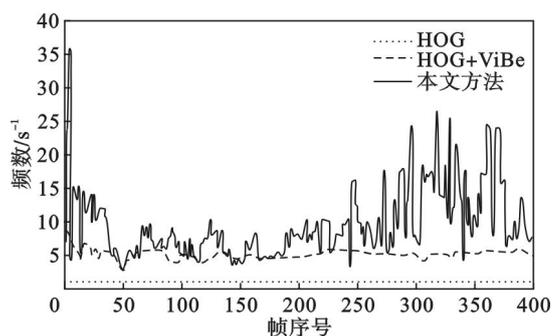


图9 实时性对比

Fig. 9 Comparison of real-time

4 结 论

本文提出了一种基于改进 ViBe 算法的行人检测方法. 该方法采用帧间差分方法找出鬼影像素, ViBe 算法的背景模型生成特征更新了背景模型, 从而有效地提高了鬼影的消除速度. 此外, 为了提高前景抽取的准确性并获得良好的算法实时性, 采用当前像素和已知像素之间的关系, 动态调整了 ViBe 算法的模型参数. 同时, 为了保证整体算法的时效性, 对 ViBe 处理得到的压缩图像的运动目标位置进行原图 ROI 的 HOG 行人检测. 在实验部分, 本文算法分别对不同大小的连续图像序列中的运动目标检测准确性和行人检测实时性加以验证. 实验结果表明本文所提出的方法能对视频中的多行人运动目标进行检测, 并且具有良好的检测准确性和实时性.

参考文献:

- [1] Brox T, Bruhn A, Papenberger N, et al. High accuracy optical flow estimation based on a theory for warping [M]. Berlin: Springer, 2004: 25 - 36.
- [2] Zhan C H, Duan X H, Xu S Y, et al. An improved moving object detection algorithm based on frame difference and edge detection [C] // Image and Graphics. Sichuan: IEEE, 2007: 519 - 523.
- [3] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction [C] // Proceedings of the 17th International Conference on Pattern Recognition. Cambridge: IEEE, 2004: 28 - 31.
- [4] Kim K, Chalidabhongse T H, Harwood D, et al. Real-time foreground-background segmentation using codebook model [J]. *Real-Time Imaging*, 2005, 11(3): 172 - 185.
- [5] Barnich O, Van Droogenbroeck M. ViBe: a universal background subtraction algorithm for video sequences [J]. *IEEE Transactions on Image Processing*, 2011, 20(6): 1709 - 1724.
- [6] Papageorgiou T D, Curtis W A, McHenry M, et al. Neurofeedback of two motor functions using supervised learning-based real-time functional magnetic resonance imaging [C] // Annual International Conference of the IEEE. Minneapolis: IEEE, 2009: 5377 - 5380.
- [7] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603 - 619.
- [8] Forssen P E, Lowe D G. Shape descriptors for maximally stable extremal regions [C] // The 11th International Conference on Computer Vision. Rio de Janeiro: IEEE, 2007: 1 - 8.
- [9] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors [J]. *International Journal of Computer Vision*, 2004, 60(1): 63 - 86.
- [10] Dalal N, Tnggs B. Histograms of oriented gradients for human detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005: 886 - 893.
- [11] 吴东飞. 静态场景下多运动目标检测与跟踪 [D]. 合肥: 合肥工业大学, 2013.
(Wu Dong-fei. Multi-moving object detection and tracking based on static scenario [D]. Hefei: Hefei University of Technology, 2013.)
- [12] Van Droogenbroeck M, Paquot O. Background subtraction: experiments and improvements for ViBe [C] // Computer Vision and Pattern Recognition Workshops (CVPRW). Providence: IEEE, 2012: 32 - 37.