

基于多尺度-多形状 HOG 特征的行人检测方法

牛 杰¹, 钱 堃²

(1. 常州信息职业技术学院 电子与电气工程学院, 江苏 常州 213164;

2. 东南大学 自动化学院, 江苏 南京 210096)

摘 要:提出一种图像中人体快速自动检测方法。提取图像的多尺度-多形状方向梯度直方图(HOG)特征向量,用于描述人体的形状特征,结合 Adaboost 机器学习法训练级联型分类器,以加速人体的检测过程。相比较传统算法,该方法没有采用静态背景模型,也不是仅仅依赖于易受外部环境因素干扰的颜色信息,从而一定程度地适应了人体姿态变化,以及非结构化环境下常见的光照波动、背景杂乱等不良因素所带来的干扰。实验验证了该方法的准确性和较高的计算效率。

关键词:方向梯度直方图;行人检测;Adaboost;机器学习

中图分类号:TP391.41

文献标识码:A

文章编号:1673-629X(2011)09-0099-04

Pedestrian Detection Based on Multi-Scale and Multi-Shape HOG Features

NIU Jie¹, QIAN Kun²

(1. Sch. of Electrical and Electronic Eng., Changzhou College of Inf. Techn., Changzhou 213164, China;

2. School of Automation, Southeast University, Nanjing 210096, China)

Abstract: A fast and automatic people detection method is proposed. The multi-scale and multi-shape histogram of oriented gradient (HOG) features are extracted, which serve as a powerful description of human shapes; The extracted features are then fed into a cascade of classifiers trained by Adaboost algorithm to greatly accelerate the people detection scheme. The proposed method is independent from background models as well as color information in images, which is highly unreliable due to disturbance. This method is robust against human posture variances, lightening fluctuations as well as background cluttering. Experimental results validate the favorable performance of high accuracy and computational efficiency of the proposed method.

Key words: histogram of oriented gradient; pedestrian detection; Adaboost; machine learning

0 引 言

人的视觉检测在智能监控^[1]、人机交互^[2]等领域都有着重要的应用价值。文中主要讨论不依赖背景模型的图像中行人检测方法,可以用于各种视角的静态/动态摄像机,是机器视觉领域中一项具有挑战性的研究课题。

基于视觉的人体检测方法可大致分为自底向上与自顶向下两类。自底向上的方法^[3,4]从图像的底层特征如边缘、颜色聚类开始分析人的肢体,并在人体结构约束下优化特定的代价函数以搜索人体存在区域。此类方法的典型代表有模板匹配法、光流法等,需要大量搜索计算,难以实时实现。自顶向下的方法^[5,6]则利

用统计学习理论训练特定目标的分类器。自适应 boosting 方法(Adaboost)^[7]不需要预先知道学习机的训练错误率,而是自动根据弱学习机的预测准确率调节相应弱学习机的权值。因此只要学习机的性能超过随机猜想且迭代充分,就能使经验错误率降至任意低的水平,并能保证推广错误率也低于某近似的上限。一般此类方法利用人体图像的基元特征,如 Harr 小波特征等构建弱分类器^[6]的级联,但对于人体目标形变的情况容易出现误检。传统的傅立叶描述子^[8]是对人体外形的全局参数化描述,忽略了局部细节形状特征,对于遮挡的鲁棒性较差。形状上下文(Shape Context, SC)^[9,10]是相对大尺度的局部描述,参考点位置可以任意选择,对虚假点和小范围的遮挡比全局法更为鲁棒。Dalal^[5]等提出了方向梯度直方图(Histogram of Oriented Gradient, HOG)描述子,是对人体姿态形变的视觉表现更为鲁棒的描述^[11]。Zhu^[12]等利用 Adaboost 法构建 HOG 特征的线性 SVM 弱分类器的级联,在检测

收稿日期:2011-03-01;修回日期:2011-06-05

基金项目:江苏省现代教育技术研究 2011 年度技术应用重点课题(2011-R-18926)

作者简介:牛 杰(1983-),男,江苏淮安人,讲师,硕士,研究领域为图像处理、测量控制。

算法运行效率上取得了显著的提升。

文中提出一种基于多尺度-多形状方向梯度直方图(HOG)特征行人快速检测跟踪算法,其特点在于:综合利用了人的视觉表观组合特征,因此有效降低了假阳率与假阴率;采用 Boosting 机器学习方法优化检测分类器,保持了算法较高运行效率;检测与跟踪算法不依赖于背景模型,因此与摄像机动静无关,具有较高的实用性。

1 多尺度-多形状 HOG 特征提取

边缘图像 I 的像素 (x, y) 对 x 和 y 方向求导:

$$d_x(x, y) = I(x+1, y) - I(x-1, y) \quad (1)$$

$$d_y(x, y) = I(x, y+1) - I(x, y-1) \quad (2)$$

(x, y) 处的梯度幅值与梯度方向角计算如下:

$$m(x, y) = \sqrt{d_x^2(x, y) + d_y^2(x, y)} \quad (3)$$

$$\theta(x, y) = \tan^{-1}(d_y(x, y)/d_x(x, y)) \quad (4)$$

在图 1 的示例中, (a) 为二值原图, (b) 为边缘图, (c) 为对应的方向梯度图, 其中一个矩形范围内的像素梯度方向以 $2\pi/9$ 为间隔向 9 个 bin 中投票, 得到该块对应中像素的 HOG, 如 (d) 图所示。

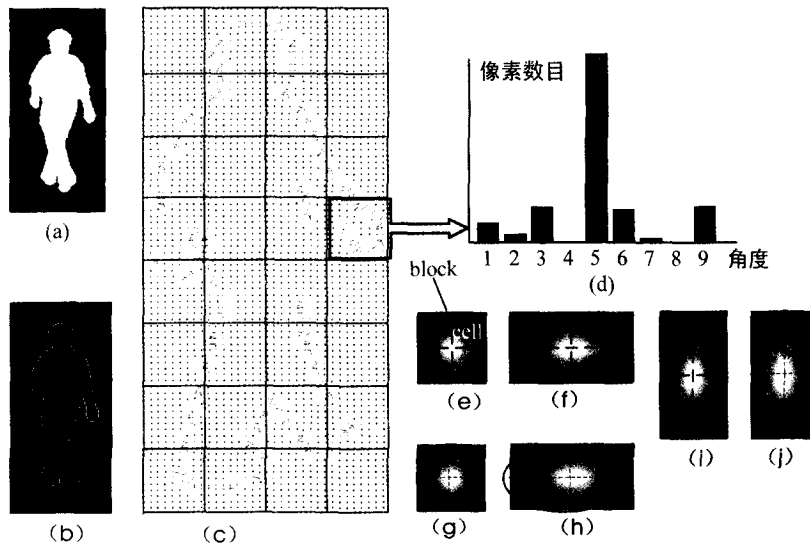


图 1 HOG 特征提取原理

Dalal^[5] 将 64×128 的检测窗口分割为若干像素为 8×8 的 cell, 每 2×2 个 cell 组成一个滑动的块 (block), 各个 block 之间相互重叠。文献[5]中指出了块在尺寸及式样上的多种变形:除了如图 1(e) 所示的 2×2 方块(R-HOG)以外,还有一种用参数 n_r 和 n_θ 分别表示径向与角度的离散化数目的圆形块(C-HOG)。但 Dalal 采用的 block 都是固定尺寸即 16×16 的, 在 64×128 的检测窗口内共 105 个, 必须评价所有这 105 个 block 才能决策出当前窗口是否属于待检目标。Zhu^[12] 认为固定尺寸的 block 在特征表达能力上存在局限性, 建议在 64×128 的窗口内采用从 12×12

到 64×128 尺寸不等的多个矩形 block, 称为 R-HOG, 并结合特征选择的机器学习算法, 证实了采用多尺度 R-HOG 能够提高检测效果。文中从形状尺度两个方面进一步丰富 block 的选型, 采用如下两种形状的多尺度 block:

1) 多尺度圆形 HOG (Circular-HOG, C-HOG)。

圆形 HOG 每个 block 仍由 2×2 的 cell 组成, 但长宽比率为 $(1:1), (1:2), (2:1)$, 尺度从 12×12 到 64×128 不等, 如图 1(g)(h)(j) 所示。C-HOG 中的投票在极坐标系下进行, 以像素 d_i 为圆心的圆形块的 HOG_{d_i} 定义在一个以 d_i 为圆心、 R 为半径、按对数距离等间隔建立的同心圆区域上:

$$h_{d_i}^o(k) = \# \{d_j \neq d_i : d_j \in E_o : (d_j - d_i) \in \text{bin}(k)\}, \quad k = 1, \dots, (n_r \cdot n_\theta) \quad (5)$$

其中符号 $\#$ 表示统计数目。点 d_i 的 SC 实际上是一个二维分布。第 k 个 bin 的径向序号与角度序号分别为 k_r 与 k_θ , 其值为该扇形范围内的像素统计数量, 是除点 d_i 以外该形状边缘上满足到 d_i 点的径向距离为 k_r , 偏角为 k_θ , 并且梯度方向为 o 的所有点的数目总和。

设点 d_j 相对参考点 d_i 所在的径向和偏角分别为 r_j 与 θ_j , 定义 d_j 的泛化 bin 域为 $S(d_j) = \{\text{bin}(l), l \in \{(\theta_d + i) \bmod n_\theta\}_{i=-b}^b\}$, 于是(5)式变为:

$$h_{d_i}^o(k) = \# \{d_j \neq d_i : d_j \in E_o : (d_j - d_i) \in S(d_j)\} \quad (6)$$

$$SC_{d_i}^o(k) = \sum_{d_j} g(\theta_d, k) \cdot h_{d_i}^o(k), \quad k = 1, \dots, (n_r \cdot n_\theta) \quad (7)$$

$g(\theta_d, k)$ 是以 θ_d 为对称中心峰值为 1 的高斯密度函数, 上述泛化考虑了以角度 θ_d 为中心的附近共 $2b + 1$ 个 bin, 并以概率加权的方式将其融合, $SC_{d_i}^o(k)$ 就是角度泛化后点

d_j 的形状上下文。

2) 多尺度矩形 HOG (Rectangular-HOG, R-HOG)。

每个 block 仍由 2×2 的 cell 组成, 但长宽比率为 $(1:1), (1:2), (2:1)$, 尺度从 12×12 到 64×128 不等, 如图 1(e)(f)(i) 所示。R-HOG 的计算方法类似于 C-HOG, 不再赘述。

HOG 弱特征与弱分类器:

每个 cell 内计算一个 9bin 的方向梯度直方图, 将每个 block 中的 4 个 cells 的直方图 bin 拉直成向量, 于是每个 block 对应一个 36 维特征向量, 并将此归一化至一个 L2 范数 (L2-norm) 单位长度。有多种范数可

选, 设 $x = [x_1, \dots, x_{36}]$ 为给定 block 中直方图的未归一化特征向量, $\|x\|_k$ 为 x 的 k -norm, $k=1, 2$, e 为一小常数 (e 的大小不影响结果), 几种常用归一化因子计算如下, 范数的选择对结果影响不大^[5], 不妨选取计算速度更快的 L2-norm:

$$\text{L2-norm: } f = \frac{x}{\sqrt{\|x\|_2^2 + e^2}}, \|x\|_2 = \sqrt{\sum_{i=1}^{36} |x_i|^2} \quad (8)$$

对于每个 block 的 36 维特征向量, 采用一个 SVM 分类器对该 block 是否属于人体进行 0/1 判决。于是一个 block 的 36 维特征向量称为弱特征, 该 SVM 二分类器称为一个弱分类器。SVM 分类器一般可采用经典的线性 SVM 或者非线性高斯核 SVM。

2 分类器级联的监督学习

Adaboost 算法^[6]用于分类的基本原理就是将简单的弱分类器通过线性组合构成复杂的强分类器:

$$f(x) = \sum_{i=1}^T \alpha_i h_i(x) \quad (9)$$

$h_i(x)$ 是弱分类器, $H(x) = \text{sign}\{f(x)\}$ 是强分类器。给定弱学习训练集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 其中 $x_i \in \chi$ 是输入训练样本向量, χ 是输入的训练样本集合, $y_i \in \{0, 1\}$ 是样本中正例和反例的分类标志, 0 表示对应反例, 1 表示对应正例, 通常反例样本数目 m 是正例样本数目 l 的数倍。初始化 K 个特征 $f_j(\cdot)$, ($j=1, \dots, K$), 某个样本 x_i 的特征就为 $f_1(x_i), \dots, f_K(x_i)$, 称为特征池。每个训练样本都被赋予一个权值, 并在迭代中调整权值。通过不断加入弱分类器, 最终达到某个预定的足够小的错误率。 T 次迭代就可以从 K 个特征的特征池中选出能够最好地描述样本特征的 T 个特征, 相应弱分类器的线性组合即强分类器。

训练 HOG 特征级联型分类器的具体步骤为, 首先为待检目标构建 64×128 尺寸的正负样本集。采用 C-HOG 与 R-HOG 这两种 block, 在 64×128 的样本图像或者检测窗口内共有 10062 个 36 维的 HOG 特征。其次采用表 1 所示的流程训练级联分类器。每一层分类器的训练只需采样一个大小为 $\log 0.05 / \log 0.95 = 59$ 的子集就能以 0.95 的概率获得最佳的 5% 的样本估计。使用 Adaboost 训练该层分类器, 直至获得最小的错误拒绝率, 若其错误拒绝率高于设定目标则引入新的层。

3 实验结果

使用文献[5]的单窗口遍历检测方法, 比较固定尺寸、单一形状(C-HOG)HOG 特征与多形状-多尺度

表 1 级联分类器训练算法

级联分类器训练算法:

初始化: 初始层数 $i = 0$, 初始检测率 $D_i = 1.0$, 初始假率 $F_i = 1.0$

F_{target} 为整体假率, f_{max} 为每层最大可接受假率, d_{min} 为最小可接受检测率

While $F_i > F_{\text{target}}$

$i = i + 1$ // 新增一层

$f_i = 1.0$

While $f_i > f_{\text{max}}$

// 计算第 i 层的误检率 f_i

(1) 使用正负样本的 500 个随机弱特征训练高斯核 SVM 分类器

(2) 使用 Adaboost 算法得到最佳 SVM 并加入强分类器, 更新该 SVM 分类器的权值

(3) 使用得到的强分类器评价正负样本

(4) 降低 thresh 直到本层的分类器检测率达到 d_{min}

(5) 计算该 thresh 下的 f_i

End

$F_{i+1} = F_i \times f_i$, $D_{i+1} = D_i \times d_i$, $\text{Neg} \leftarrow \varphi$

If $F_i > F_{\text{target}}$, 用当前的级联分类器测试所有负样本, 将分类错误的样本归入 Neg 集

End If

End

HOG 特征的性能, 后者算法中, 多尺度-多形状的相关参数 n_r 、 n_θ 以及 SVM 验证步骤需要确定的阈值 η 等均在实验中人为设定。改变图像中遍历的窗口数目从 800 到 12800 共 10 种情况, 得到不同的检测性能, 用 ROC (Receiver Operating Characteristic) 曲线表示, 描述了误检率 (False Positive Rate, FPR) 与正确检测率 (True Positive Rate, TPR) 之间的变换关系。如图 2 所示, 文中多形状-多尺度 HOG 特征 ROC 的曲线与横轴间所围成的面积较大, 表示在所有代价比上的平均性能较优, 也反映了多尺度-多形状 HOG 特征描述正反两类样本的可分性好于固定尺寸、单一形状的 HOG 特征。实验中, 当图像中遍历的窗口数目达到 12800 时正确检测率最高达到 91.5%, 但单帧耗时达到 5s。

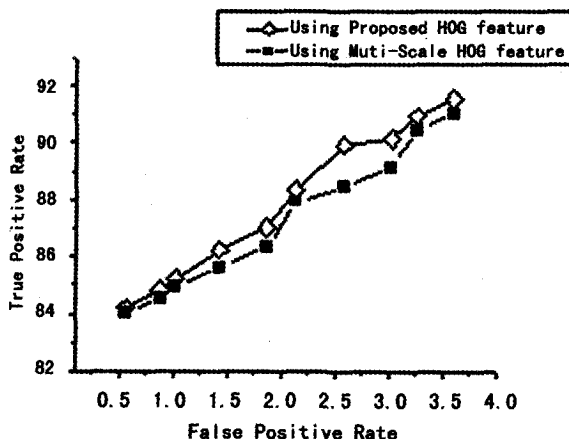


图 2 变尺度 HOG 方法与文中方法的 ROC 曲线比较



图 3 图像中人体检测结果

多尺度-多形状 HOG 特征分类器级联在自建行人图像库 IDIR-human 的检测结果如图 3 所示,可见对于单人无遮挡情况一般具有较高检测率。当人体受部分遮挡时出现了一些漏检现象,对于某些线条错综的背景区域出现了误检。选用文献[5,6]这两种主流的准实时/实时人体检测算法与本方法进行比较。表 2 的比较结果可见,本方法可获得与文献[5]方法接近的检测率,但显著加快了运行速度;相比文献[6]方法则明显提高了正确检测率。实验均为在 Intel 双核 2.66GHz PC 机上对 320×240 分辨率图像检测单人所得。

4 结束语

提出一种基于视觉的快速目标人检测算法,提取图像的多尺度-多形状方向梯度直方图(HOG)特征,采用 Adaboost 算法训练分类器级联,检测结果对自然光照条件变化、背景杂乱、人体姿态变化及部分遮挡等情况具有一定适应性。相比传统检测方法,实验验证了该方法具有较高的检测率和良好的计算实时性。

表 2 人体检测正确率比较

目标	文献[5]方法			文献[6]方法			本方法		
	正确率%	漏,误检率%	算法耗时 ms	正确率%	漏,误检率%	算法耗时 ms	正确率%	漏,误检率%	算法耗时 ms
正/背面	91.0	5.4, 3.6	1214.2±1568.24	56.3	35.5, 8.2	20.5±65.3	90.1	6.3, 3.6	35.7±38.54
侧面	88.9	7.0, 4.1	1214.2±1568.24	51.8	37.1, 11.1	30.2±89.1	87.8	7.6, 4.6	51.3±48.34

参考文献:

- [1] Siebel N T. Design and Implementation of People Tracking Algorithms for Visual Surveillance Applications[D]. UK: The University of Reading, 2003.
- [2] Shao X, Zhao H, Nakamura K, et al. Detection and Tracking of Multiple Pedestrians by Using Laser Range Scanners[C]// Proc of IEEE International Conference on Intelligent Robots and Systems. San Diego, CA, USA: [s. n.], 2007: 2174-2179.
- [3] Srinivasan P, Shi J B. Bottom-up recognition and parsing of the human body[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: [s. n.], 2007: 1-8.
- [4] Ren X, Berg A C, Malik J. Recovering human body configurations using pairwise constraints between parts[C]// IEEE International Conference on Computer Vision. Beijing, China: [s. n.], 2005: 824-831.
- [5] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). [s. l.]: [s. n.], 2005: 886-893.
- [6] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [s. l.]: [s. n.], 2001.
- [7] Freund Y, Schapire R E. Experiments with a new boosting algorithm[C]// Proceedings of the 13th Conference on Machine Learning. Bari, Italy: Morgan Kaufmann, 1996: 148-156.
- [8] Mowbray S D, Nixon M S. Automatic Gait Recognition via Fourier Descriptors of Deformable Objects[C]// Proc of the 4th International Conference on Audio and Video Based Biometric Person Authentication. Guildford, UK: [s. n.], 2003: 566-573.
- [9] Mori G, Belongie S, Malik J. Efficient Shape Matching Using Shape Contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(11): 1832-1837.
- [10] Liu D, Chen T. Soft shape context for interactive closest point (下转第 106 页)

分割的子段作为分割对象,避免了问题陷入局部最优的困境。通过实验证明,与同为固定分段数的时间序列分段算法 PLR_PAA 方法比较,虽然计算耗时略高,但是 PLR_FPIP 算法的拟合误差也相对较小,拟合精度更高。

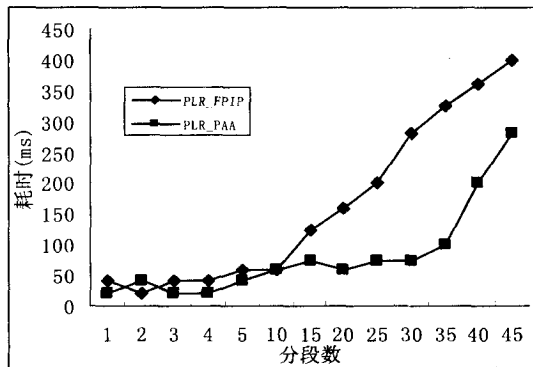


图 4 算法耗时比较

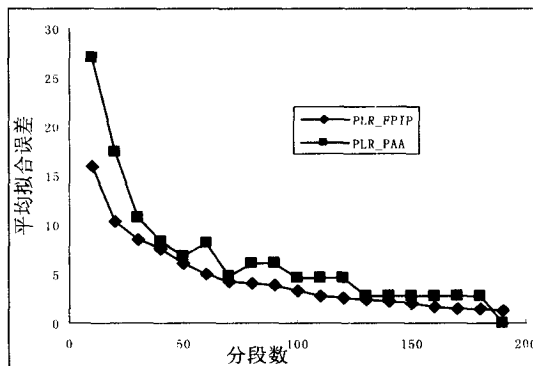


图 5 拟合误差比较

参考文献:

- [1] 张 军, 马志民. 基于时间序列的相似子模式发现算法[J]. 计算机技术与发展, 2006, 16(1): 140-141.
- [2] Keogh E, Chakrabarti K, Pazzani M J, et al. Dimensionality reduction for fast similarity search in large time series databases[J]. Knowledge and Information Systems, 2008, 3(3): 263-286.
- [3] Keogh E, Kasetty S. On the need for time series data mining benchmarks: A survey and empirical demonstration[C]//Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [s. l.]: [s. n.], 2002: 102-111.
- [4] Tewari G, Snyder J, Sander P V, et al. Signal-specialized parameterization for piecewise linear reconstruction[C]//Eurographics Symposium on Geometry Processing. [s. l.]: [s. n.], 2004: 39-52.
- [5] Yi B K, Faloutsos C. Fast time sequence indexing for arbitrary Lp norms[C]//Proc. of the 26th International Conference on Very Large Databases. San Francisco: Morgan Kaufmann Publishers Inc, 2002: 385-394.
- [6] Keogh E, Chu S, Hart D, et al. Segmenting time-series: A survey and novel approach[C]//Data Mining in Time-series Databases. [s. l.]: World Scientific, 2004: 1-22.
- [7] Keogh E, Chakrabarti K, Pazzani M J, et al. Locally adaptive dimensionality reduction for indexing larger time series databases[C]//Proc. of ACM SIGMOD Conference on Management of Data. Santa Barbara, CA: [s. n.], 2001: 151-162.
- [8] Park S, Kim S W, Cho J S, et al. Prefix querying: an approach for effective subsequence matching under time warping in sequence databases[C]//Proceedings of the 10th International Conference on Information and Knowledge Management. New York: ACM Press, 2002: 255-262.
- [9] Lavrenko V, Schmill M, Lawrie D, et al. Mining of concurrent text and time series[C]//Proceedings of the 6th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining Workshop on Text Mining. Boston, MA: [s. n.], 2002: 37-44.
- [10] 孙焕良, 邱邦华, 魏溯华. 一种优化的自底向上时间序列分段算法[J]. 沈阳建筑大学学报自然科学版, 2007, 23(6): 1050-1051.
- [11] 肖 辉, 马海兵, 龚 薇. 基于时态边缘算子的时间序列分段线性表示[J]. 计算机工程与应用, 2008, 44(19): 156-159.
- [12] 詹艳艳, 徐荣聪, 陈晓云. 基于斜率提取边缘点的时间序列分段线性表示方法[J]. 计算机科学, 2006, 33(11): 139-142.
- [13] 杜 奕, 卢德唐, 李道伦, 等. 基于层次聚类的时间序列在线划分算法[J]. 模式识别与人工智能, 2007, 20(3): 23-27.
- [14] Chung F L, Fu T C, Luk R, et al. Time series pattern matching based on perceptually important points[C]//Proc. of the IJCAI Workshop on Learning from Temporal and Spatial Data. [s. l.]: [s. n.], 2001: 1-7.
- [15] 周大镛, 李敏强. 基于序列重要点的时间序列分割[J]. 计算机工程, 2008, 34(23): 15-16.
- [16] registration[C]//International Conference on Image Processing. Singapore: [s. n.], 2004: 1081-1084.
- [17] 周 柯. 基于 HOG 特征的图像人体检测技术的研究与实现[D]. 武汉: 华中科技大学, 2008.
- [18] Zhu Q, Avidan S, Yeh M C, et al. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients[C]//IEEE Conference on Computer Vision and Pattern Recognition. [s. l.]: [s. n.], 2006: 1491-1498.

(上接第 102 页)