

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2012 3
Vol.17 No.

中国科学院遥感应用研究所
中国图象图形学学会主办
北京应用物理与计算数学研究所

中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2012 年 3 月 第 17 卷 第 3 期(总第 191 期)

目 次

图像处理 and 编码

摄影测量共线方程的单位四元数描述	杨化超,卢晓攀,王永波,姚国标(301)
分形压缩感知高维信号重构方法	刘佑鑫,孙权森(309)
自适应子矢量划分的快速码字搜索算法	吴鑫鹏,潘志斌,李达(315)
基于人类视觉模型的各向异性扩散滤波方法	宋建军,侯志强,余旺盛(321)
整数提升小波多相矩阵分解系数的快速提取算法	王锋,赵志文,牟盛(329)
自适应样本块大小的图像修复方法	孟春芝,何凯,焦青兰(337)

图像分析和识别

改进 FCM 在交互式图像分割中的应用	依玉峰,高立群,郭丽(342)
输电导线图像目标识别方法	孙凤杰,杨镇潭,李媛媛,范杰清(349)
无表观特征小目标检测与跟踪	陶霖密,李亮,邸慧军(357)
应用改进频率调谐的海上小目标检测方法	任蕾,施朝健,冉鑫(365)
视觉注意原理局部特征的行人检测	刘俊涛,刘文予,吴彩华,李雄伟,冯镔(370)
尺度不变单样本人脸识别方法	王炜强,张晓阳,曹春芹,付奎生(380)
复杂交通场景中采用稀疏表示的车辆识别方法	李修志,吴健,崔志明,陈建明(387)
利用复合导数的边缘检测新算法	程金梅,叶永强,姜斌(393)

图像理解和计算机视觉

迭代分割逼近:新的基于局部响应显著度的角点检测	马丽红,谭幸均(402)
融合 SIFT 特征的熵图估计医学图像非刚性配准	张少敏,支力佳,赵大哲,林树宽,赵宏(412)

计算机图形学

面向非均匀采样点集的 3 维表面重建算法	刘晓平,段瑞青,余烨(419)
基于可编程图形管线的大规模流程工厂模型多分辨率绘制方法	周剑,唐卫清,朱耀琴,夏明,黄晓剑(426)
带有位置修正的环境映射	王晨昊,汤晓安,孙即祥,马伯宁(435)

遥感图像处理

海岸线遥感信息提取的元胞自动机方法及其应用	冯永玖,韩震(441)
-----------------------------	-------------

中国图象图形学报

刊名题字: 宋 健	月刊(1996 年创刊)	第 17 卷 第 3 期	2012 年 3 月 16 日出版
-----------	--------------	--------------	-------------------

主管单位	中国科学院	Superintended by	Chinese Academy of Sciences
主 办	中国科学院遥感应应用研究所 中国图象图形学学会 北京应用物理与计算数学研究所	Sponsored by	Institute of Remote Sensing Application, CAS China Society of Image and Graphics Institute of Applied Physics and Computational Mathematics
主 编	李小文	Chief editor	LI Xiaowen
编辑出版	《中国图象图形学报》编辑出版委员会 北京 9718 信箱 邮编 100101 电子信箱:jig@irsa. ac. cn 电话:010-68407995 010-82614429 网 址:www. cjig. cn	Editor ,Publisher	Editorial and Publishing Board of Journal of Image and Graphics (P. O. Box 9718 ,Beijing 100101 ,China) E-mail:jig@irsa. ac. cn
印刷装订	北京北林印刷厂	Distributed by	Beijing Bureau for Distribution of Newspapers and Journals
广告经营许可证	京朝工商广字第 0346 号	Domestic	All Local Post Offices in China
总 发 行	北京报刊发行局	Foreign	China International Book Trading Corporation (P. O. Box 399 ,Beijing 100044 ,China)
订 购	全国各地邮局	Printed by	Beijing Beilin Printing House
国外发行	中国国际图书贸易总公司 (中国国际书店) (北京 399 信箱 邮编 100044)		

Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 17 No. 3 March 2012

Contents

Image Processing and Coding

Unit quaternion based description of collinearity equations	Yang Huachao, Lu Xiaopan, Wang Yongbo, Yao Guobiao (301)
Fractal compressive sensing for high-dimension signal recovery	Liu Jixin, Sun Quansen (309)
Fast codeword search algorithm based on adaptive subvector partitions	Wu Xinpeng, Pan Zhibin, Li Da (315)
Anisotropic diffusion image smoothing method based on human visual model	Song Jianjun, Hou Zhiqiang, Yu Wangsheng (321)
Fast extraction algorithm of the polyphase matrix decomposition coefficient based integer lifting wavelet	Wang Feng, Zhao Zhiwen, Mou Sheng (329)
Image completion method with adaptive patch size	Meng Chunzhi, He Kai, Jiao Qinglan (337)

Image Analysis and Recognition

Application of improved FCM for interactive image segmentation	Yi Yufeng, Gao Liquan, Guo Li (342)
Methods of transmission line target recognition	Sun Fengjie, Yang Zhenhuan, Li Yuanyuan, Fan Jieqing (349)
Featureless small object detection and tracking	Tao Linmi, Li Liang, Di Huijun (357)
Small target detection method under sea surface environment using revised frequency tuned saliency detection	Ren Lei, Shi Chaojian, Ran Xin (365)
Pedestrian detection method using local feature based on vision attention	Liu Juntao, Liu Wenyu, Wu Caihua, Li Xiongwei, Feng Bin (370)
Scale invariant face recognition from single sample	Wang Weiqiang, Zhang Xiaoyang, Cao Chunqin, Fu Kuisheng (380)
Sparse representation method of vehicle recognition in complex traffic scenes	Li Xiuzhi, Wu Jian, Cui Zhiming, Chen Jianming (387)
Novel edge detection algorithm using a composite derivative	Cheng Jinmei, Ye Yongqiang, Jiang Bin (393)

Image Understanding and Computer Vision

Iterative segment approaching: new corner detection based on local response saliency	Ma Lihong, Tan Xingjun (402)
Entropic graph estimation integrated with SIFT features for medical image non-rigid registration	Zhang Shaomin, Zhi Lijia, Zhao Dazhe, Lin Shukuan, Zhao Hong (412)

Computer Graphics

Three-dimensional surface reconstruction algorithm for non-uniform sampling points	Liu Xiaoping, Duan Ruiqing, Yu Ye (419)
Multi-resolution rendering approach of large-scale process plant models based on programmable graphics pipeline	Zhou Jian, Tang Weiqing, Zhu Yaoqin, Xia Ming, Huang Xiaojian (426)
Environment mapping with position rectification	Wang Chenhao, Tang Xiao'an, Sun Jixiang, Ma Boning (435)

Remote Sensing Image Processing

Cellular automata approach to extract shoreline from remote sensing imageries and its application	Feng Yongjiu, Han Zhen (441)
---	------------------------------

中图分类号: TP391.4 文献标志码: A 文章编号: 1006-8961(2012)03-0370-10

论文引用格式: 刘俊涛, 刘文予, 吴彩华, 李雄伟, 冯滨. 视觉注意原理局部特征的行人检测[J]. 中国图象图形学报, 2012, 17(3): 370-379

视觉注意原理局部特征的行人检测

刘俊涛^{1,2}, 刘文予¹, 吴彩华³, 李雄伟², 冯滨¹

1. 华中科技大学电子与信息工程系, 武汉 430074;

2. 机械工程学院计算机工程系, 石家庄 050003;

3. 空军雷达学院信息对抗系, 武汉 430019

摘要: 在复杂背景下检测行人, 具有重要的理论和应用价值。为了适应此类场景中光照的变化和行人姿态的多样性, 依据人眼视觉注意原理, 提出基于视觉注意的局部特征。该特征具有光照和旋转不变性, 并能用于多尺度分析。采用基于特征块的行人表示模型, 行人被表示为特征块的集合。每一个特征块用基于视觉注意局部特征的统计直方图和位置关系表示。用聚类的方法得到基于特征块的行人模型。依据每一个特征块在检测窗口中的最大响应训练 AdBoost 检测分类器, 并用困难负样本和可信样本提高检测分类器的性能。用滑动窗口方法在图像和尺度空间中找到检测分类器的局部最大响应, 以确定行人位置。实验结果表明, 与现有方法相比, 本文方法对竖直边缘不敏感, 可以处理一定程度的遮挡以及姿态变化。

关键词: 视觉注意; 局部特征; 行人检测; 特征块

Pedestrian detection method using local feature based on vision attention

Liu Juntao^{1,2}, Liu Wenyu¹, Wu Caihua³, Li Xiongwei², Feng Bin¹

1. Department of Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074, China;

2. Department of Computer Engineering, Mechanical Engineering Institute, Shijiazhuang 050003, China;

3. Information Countermeasure Department, Air Force Radar Academy, Wuhan 430019, China

Abstract: Pedestrian detection in images with complex backgrounds is valuable in theory and applications. To deal with the variability of illumination and pedestrian's poses, the local feature based on vision attention (LFVA), derived from a visual saliency mechanism, is proposed in this paper. LFVA is illumination and rotation invariant, and can also be used for multi-scale analysis. A pedestrian model based on feature blocks is proposed, in which the pedestrians are represented by a set of feature blocks. Every feature block is represented by the position and histogram of the LFVA. The pedestrian model is obtained by clustering. The AdBoost detection classifier is trained using the maximal responses of the feature blocks and is improved using hard negative samples and trusted samples. The local maximal of the response of the detection classifier in image and scale space is located as the position of the pedestrian by sliding window search. Compared with the existed methods, the proposed method is less sensitive to vertical edges and can deal with occlusion and pose variation to some extent.

Key words: vision attention; local feature; pedestrian detection; feature block

收稿日期: 2011-04-08; 修回日期: 2011-08-15

基金项目: 国家自然科学基金项目(60903172); 河北省自然科学基金项目(F2009001435)

第一作者简介: 刘俊涛(1979—), 男, 讲师, 现为华中科技大学通信与信息系统专业博士研究生, 主要研究方向为计算机视觉、模式识别。E-mail: ropobox@21cn.com

0 引言

近年来,在图像中检测行人备受人们关注,该技术可用于智能监控^[1]、智能交通^[2]和运动分析^[3]等领域。本文关注如何在复杂场景的单幅图像中检测行人。在此类场景中,光照时常发生变化;景物复杂多样;行人具有多变的服饰和姿态;行人与行人之间、行人与景物之间常常相互遮挡;行人图像分辨率往往较低,只占据了视野中的一小部分。在此种复杂环境中准确的检测行人异常困难,而这又是智能监控和智能交通中常见的情形,具有重要的研究和应用价值。

在给定的图像 I 中检测行人可以归结为在尺度和图像空间中计算检测窗口出现行人的概率 $P(W_{x,y,l})$, 其中 $W_{x,y,l}$ 表示尺度为 l 、坐标为 (x,y) 处的检测窗口。在计算 $P(W_{x,y,l})$ 时需要按照行人的表示方法提取检测窗口 $W_{x,y,l}$ 中图像的特征。根据行人表示方法的不同,现有的行人检测方法大致可以分为两类:一类不使用模型,仅用一些底层的特征来表示行人,称为无模型方法^[3-6];另一类方法用事先设计好的模型表示行人,称为有模型方法^[3-5]。

HoG 方法^[6]是典型的无模型方法。该方法首先把每一个检测窗口划分为 8×8 像素的小块 (cell), 每 2×2 个小块组成一个块 (block), 相邻块之间有部分重叠。然后计算每个小块梯度方向的统计直方图 (HoG), 并连接在一起构成表示块特征的向量。最后,把检测窗口中所有块的特征向量输入到一个 SVM 检测分类器中,以确定检测窗口是否有行人。HoG 方法在分辨率较高情况下有很好的检测效果,但检测速度较慢,难以用于需要实时处理的场合^[7]。文献[8]利用基于类 Haar 小波的特征在视频的相邻两帧中检测行人。该方法利用了运动信息和行人的外观信息。级联的检测分类器提供了接近于实时的检测速度^[7]。文献[9]提出一种基于多种视觉特征的实时检测和追踪行人的系统。该方法利用基于形状和纹理的方法在感兴趣的区域 (ROI) 中检测目标。文献[10]利用人眼的视觉注意原理,提出了用视觉注意特征 (VFS) 来表示、检测行人。需要指出的是文献[9-10]中的方法都需要利用双目视觉提取感兴趣的区域或目标,然后在其中进行分类检测。

近年来出现了一些有模型的行人检测方法。例

如文献[5]将行人表示为部分 (part) 的集合,用轮廓段表示其特征。该方法在训练样本中随机取一些轮廓段,并用对称的 Chamfer 距离聚类,构成 Code Book。在检测窗口中计算每一部分的响应,并将其输入到 Boost 检测分类器中,以判断该窗口中是否有行人。该方法采用顺序扫描的方法,结合 Mean Shift 找到图像尺度空间的局部极大点,确定行人的位置。文献[3]使用星形的模型表示行人。行人的每一部分用滤波器、相对于根节点的位置、变形代价来表示,并与中心点相连。文献[3]的检测分类器是 latent SVM。文献[4]用一种新的模型——骨架联合树表示一类目标,该模型由训练样本的骨架树学习得到,保留了目标的主要结构特征而去除了目标的细节特征。该方法用匹配边界的方法快速检测目标骨架联合树中的某一支,并据此检测整个目标。

总的来看,现有的行人检测方法大多对较强的竖直边缘比较敏感^[7],这虽然是行人的一个显著特征,但也是背景中大多数景物 (如树木、建筑等) 的特征。如果不对竖直边缘进行区分,容易造成误检,事实上多数误检都是在这种位置产生的。为解决此问题,采用基于特征块的模型表示行人,此种模型并不把行人看做一个整体,而是把行人表示为一些特征块的集合。与文献[3]的不同之处在于,行人模型中的特征块大小固定、面积相对于整个行人比较小,不存在较长的竖直边缘,因此本文的方法能改善对竖直边缘的敏感性。同时,行人模型的特征块之间没有固定的相对位置关系,更具有灵活性。此外,行人模型能方便的使用部分匹配技术,即使在行人被部分遮挡的情况下仍能找到未被遮挡的特征块,可以提高检测率。现有方法对低分辨率的行人检测效果并不理想。对此,提出基于视觉注意的局部特征 (LFVA) 来表示特征块。该特征利用了人眼的视觉注意原理,可以通过改变尺度来适应不同的分辨率,同时具有光照和旋转不变性。用 INRIAPerson 数据集验证提出的方法,实验取得了较好的效果。

1 新的行人检测方法

1.1 基本思想

图1显示了本文方法的框架。方法分为训练和检测两个阶段。在这两个阶段中,都需要提取图像基于视觉注意的局部特征 (LFVA), 该特征以人眼

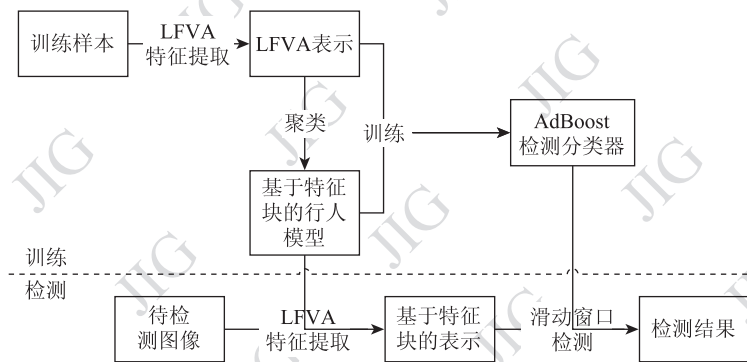


图 1 本文方法框架

Fig. 1 Framework of the proposed method

的视觉注意原理为基础,具有光照和旋转不变的特性。LFVA 的多尺度特性能保证本文的方法在低分辨率下也有较好的表现。LFVA 在 1.2 节中详细介绍。

把行人表示为特征块的集合,每一个特征块代表了行人具有显著特征的区域,用该区域相对行人包围盒的位置和 LFVA 直方图来表示特征块。这种行人模型既能处理因行人关节运动产生的姿态变化以及部分遮挡的情况,又能降低对图像中竖直边缘的敏感性,有利于提高检测的性能。行人的表示模型在 1.3 节中详细介绍。

在训练阶段,从训练图像的行人包围盒中随机取若干矩形区域,并对这些矩形区域聚类,得到基于特征块的行人表示模型。为了保证模型的一般性,采用了分层聚类的方法。然后,根据行人模型,训练 AdBoost 检测分类器,并在训练数据集和测试数据集中挖掘,以提高该分类器的性能。行人模型和检测分类器的训练在 1.4 节中详细介绍。

在检测阶段,用滑动窗口方法在图像和尺度空间中计算行人出现的概率,该概率的局部极大值处即为行人出现的位置。

1.2 基于视觉注意的局部特征及其性质

根据文献[11]人眼视网膜中的视觉细胞由两个区域构成:中心区域和周围区域。根据敏感方式的不同,视觉细胞可分为两种,一种是向心(on-center)视觉细胞,对中心明亮、周围黑暗的区域敏感;另一种视觉细胞则相反,对中心黑暗、周围明亮的区域敏感,称为离心(off-center)视觉细胞。由于圆形区域的计算比较复杂,通常可以简化为比较两个矩形区域亮度的差异。

大部分视觉注意系统,例如 iNVT^[12-13]、

VOCUS^[14]、VSF^[10],都是以此为基础,用像素的灰度之和表示亮度,通过计算中心区域与周围区域亮度之差得到该处引起人眼视觉注意的程度,把注意程度较高的区域标记为感兴趣的区域(ROI),认为这些区域吸引了人眼的大部分注意力^[15]。事实上,这些视觉系统旨在利用视觉注意原理在全局上得到 ROI,而不考虑目标是否在 ROI 中出现。而与全局信息相比,目标的局部信息在人眼对目标的检测和识别中往往占据着更重要的地位。例如,人眼只需观察物体的局部就能识别该物体。正是基于这一原理,提出基于视觉注意的局部特征。该特征在一个较小的范围内,计算图像的视觉注意特性,并据此表示目标。

如果用内圆周上像素点的灰度近似表示内圆区域的亮度,用外圆周上像素点的灰度近似表示内圆周与外圆周之间区域的亮度,可以方便地得到向心差异和离心差异。据此,提出基于视觉注意的局部特征(LFVA),即通过比较以中心像素为圆心的两个同心圆周上对应采样点的灰度差异得到 LFVA 的每一个二进制位。在灰度图像 I 中的某个像素点 (x, y) 处的 LFVA 为

$$\text{LFVA}_{r_1, r_2, n}(x, y) = \sum_{i=0}^{n-1} S(I(x_{r_1, i}, y_{r_1, i}) - I(x_{r_2, i}, y_{r_2, i})) \cdot 2^i \quad (1)$$

式中, n 是圆周上采样点的个数, r_1, r_2 分别是两个圆周的半径。 $(x_{r, i}, y_{r, i})$ 表示半径为 r 的圆周上第 i 个采样点坐标。采样点均匀分布在圆周上,采样点的坐标为

$$\begin{cases} x_{r, i} = x + r \cos(\alpha) \\ y_{r, i} = y + r \sin(\alpha) \end{cases} \quad (2)$$

$$\alpha = 2\pi/n; \quad i = 0, 1, \dots, n-1;$$

$I(x, y)$ 表示点 (x, y) 处图像的灰度值, 如果 x, y 不是整数则采用插值的方法计算 $I(x, y)$ 。为了减少计算量, 也可以对 x, y 进行取整, 得到近似的灰度值。在式(1)中, $S(x)$ 是符号函数, 当且仅当 $x > 0$ 时 $S(x) = 1$, 否则 $S(x) = 0$ 。

式(1)需要计算圆周上点的坐标, 计算复杂度较高, 也可以用正方形边界上的采样点做近似, 称为 R-LFVA。当 r_1, r_2 较小时, 误差可以忽略。当 $r_1 = 1, r_2 = 0, n = 8$ 时, R-LFVA_{1,0,8} 即为 LBP^[16]。

如前所述, 如果用内圆周上像素点的灰度近似表示内圆区域的亮度, 用外圆周上像素点的灰度近似表示内圆周与外圆周之间区域的亮度, 可以认为 LFVA 计算了局部的视觉注意特征。即, 如果 $r_1 > r_2$, LFVA 计算了外圆周亮度与内圆周亮度的差, 表示的是离心差异模式; 如果 $r_1 < r_2$, LFVA 计算了内圆周亮度与外圆周亮度的差异, 表示的是向心差异模式。随着圆周上采样点数量 n 的增加, 这种表示将越来越准确。改变 r_1, r_2 的数值可以得到不同尺度下的 LFVA 特征。由图 2 可以看到, 随着尺度增加, LFVA 的细节显示出由精细到粗略的变化过程。LFVA 的多尺度特性使其在低分辨率下也有较强的表现能力。

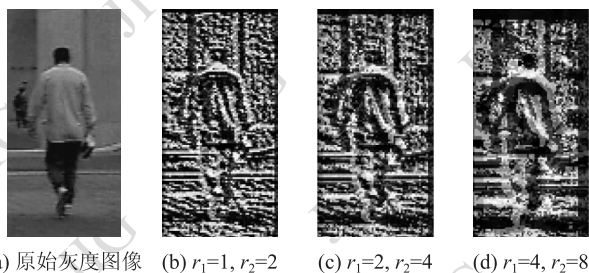


图2 原图及其不同尺度下的 LFVA

Fig. 2 Original image and it's LFVA with difference scales

LFVA 具有光照不变性。设 I, I' 分别表示光照变化前后的灰度图像。假设光照的变化在一定区域内满足一致性, 即 $\exists \delta, \forall (x, y) \in \delta, I(x, y) - I'(x, y) = C, C$ 为常数, 则对任意的 $(x_1, y_1), (x_2, y_2) \in \delta$ 有 $I(x_1, y_1) - I(x_2, y_2) = I'(x_1, y_1) - I'(x_2, y_2)$, 所以, 式(1)计算的 LFVA 在光照变化前后保持不变。

为了获得旋转不变的特性, 我们采用类似于文献[16]中的方法, 把 LFVA 值做 n 次循环移位, 每次

移 1 位, 取其最小值, 计算方法如下:

$$RI-LFVA_{r_1, r_2, n}(x, y) = \min_{i=0}^{n-1} cyc_lshift_n(LFVA_{r_1, r_2, n}(x, y), i) \quad (3)$$

式中, $cyc_lshift_n(x, i)$ 表示对 n 位的无符号数 x 循环左移 i 位(循环右移操作也可以得到相同的结果)。RI-LFVA 对旋转 $\alpha = 2\pi/n$ 的整数倍角是不变的, 当 n 趋向于无穷时对任意的旋转角都是不变的^[16]。

LFVA 的光照不变和旋转不变的特性给复杂场景下目标的表示和识别带来了方便。

1.3 行人的表示

采用基于特征块的行人表示方法, 把行人看做是特征块的集合。特征块表示了行人的某个明显特征, 对应于行人包围盒中的一个矩形区域。图 3 是行人模型的示意图, 红色的矩形区域为行人的包围盒, 蓝色的矩形区域为行人的特征块, 这里仅标出了部分特征块。每一个特征块用 LFVA 直方图(HoL)和其相对于行人包围盒的位置来表示。这种表示模型, 并不把行人作为一个整体来处理, 分割了行人较长的竖直边缘, 不会如现有方法一样对较强的竖直边缘敏感。基于特征块的模型可以处理因行人关节运动造成的形变, 即当行人姿态发生变化时, 特征块不会发生剧烈变化, 这样能够提高检测率。此种方法还可以处理部分遮挡的情况, 即行人部分肢体被遮挡时, 仍然可以找到未被遮挡部分的特征块, 有利于降低漏检率。下面分别详细介绍特征块和基于特征块的行人模型。

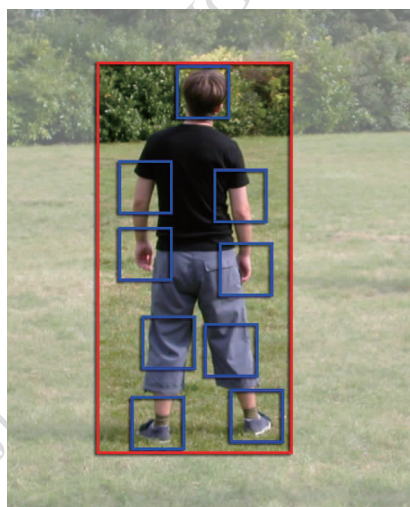


图3 行人模型示意图

Fig. 3 Pedestrian model

1.3.1 特征块的表示

特征块是行人包围盒中的一个大小为 $h_f \times w_f$ 的矩形区域。特征块 F 可以表示为一个二元组 $F = (V, P)$, 其中 $P = (x, y)$ 是特征块左上角的坐标。在这里, 坐标系以包围盒的左上角点为原点。向量 V 是特征块的 LFVA 直方图 (HoL)。HoL 的计算方法是首先把特征块分割为若干大小相等的子块, 每一个子块之间有一定区域的重叠。然后计算子块区域中每一个像素的 LFVA, 并计算直方图。最后按照从左到右、从上到下的顺序连接子块直方图, 式(4)表示了 HoL 的计算方法:

$$\text{HoL} = \oplus_i \text{Hist}_m(\text{LFVA}(x, y) | (x, y) \in B_i) \quad (4)$$

式中, $\oplus$ 表示向量的连接操作, $\text{Hist}_m(\cdot)$ 表示有 m 个统计区间 (bin) 的直方图, B_i 表示第 i 个子块。通常, 子块的长和宽分别是特征块长和宽的 $1/2$, 重叠区域的长和宽分别是子块长和宽的 $1/2$ 。式(4)中的 LFVA 算子也可替换为 R-LFVA 或 RI-LFVA, 以提高计算效率或获得旋转不变特性。

1.3.2 行人模型

行人模型为 $M = \{F_i\}$, 其中 $F_i = (V_i, P_i)$ 是第 i 个特征块, V_i 是第 i 个特征块的 LFVA 直方图 (HoL), $P_i = (x_i, y_i)$ 是第 i 个特征块相对于行人包围盒的位置。

设 $W_{x,y,l}$ 是尺度为 l 、左上角坐标为 (x, y) 的检测窗口, 其大小为 $h_w \times w_w$ 。用 $P((x', y') | F_i)$ 表示特征块 F_i 出现在 $(x', y') \in W_{x,y,l}$ 处的似然概率, 计算 $P((x', y') | F_i)$ 除要考虑 HoL 差异外, 还要考虑位置的差异, 定义 $P((x', y') | F_i)$ 为 HoL 差异与位置差异的加权和:

$$P((x', y') | F_i) = b_0 \exp(-D(V(x', y'), V_i)) + b_1 \exp(-D((x', y'), P_i)) \quad (5)$$

式中, $D(u, v)$ 表示两个向量 u, v 的欧氏距离; $V(x, y)$ 表示左上角在点 (x, y) 处的矩形区域 (与特征块 F_i 的大小一致) 的 HoL, 用式(5)计算; $b_0, b_1 \in [0, 1]$, $b_0 + b_1 = 1$ 是常数, b_0, b_1 分别表示了 HoL 差异和位置差异的权重。用 $P(W_{x,y,l} | F_i)$ 表示特征块 F_i 出现在检测窗口 $W_{x,y,l}$ 中的似然概率, $P(W_{x,y,l} | F_i)$ 定义为式(5)在检测窗口中的最大值, 即:

$$P(W_{x,y,l} | F_i) = \max_{(x', y') \in W_{x,y,l}} \{P((x', y') | F_i)\} \quad (6)$$

检测分类器 C 是一个 AdBoost 分类器, 它依据行人模型 $M = \{F_i\}$, 判定检测窗口 $W_{x,y,l}$ 中是否有行人出现。由于每一个特征块表示了行人的一个特

征, 对每一个特征训练一个弱分类器, 便于 AdBoost 分类器自动选择最显著的特征块, 同时也可实现部分匹配。

$$C(W_{x,y,l}, M) = \sum_i a_i C_w(P(W_{x,y,l} | F_i) > \theta_i) \quad (7)$$

式中, a_i 是第 i 个弱分类器的权重, $C_w(p > \theta_i)$ 为第 i 个特征块对应的弱分类器:

$$C_w(p > \theta_i) = \begin{cases} 1 & p > \theta_i \\ -1 & p \leq \theta_i \end{cases} \quad (8)$$

用式(9)定义检测窗口 $W_{x,y,l}$ 中出现行人的概率 $P(W_{x,y,l} | M)$, 其中检测分类器的输出被变换到 $[0, 1]$ 区间, 检测分类器的输出数值越大行人在窗口中出现的概率越大。

$$P(W_{x,y,l} | M) = \frac{1}{1 + \exp(-C(W_{x,y,l}, M))} \quad (9)$$

下面将详细介绍行人模型、检测分类器的训练和行人检测方法。

1.4 训练

训练分为两个阶段, 首先需要训练基于特征块的行人模型, 即要得到模型中各个特征块的位置和 LFVA 直方图。然后, 根据行人表示模型, 训练 AdBoost 检测分类器。检测分类器能够判断一个固定大小的检测窗口中是否有行人出现以及出现的概率。为了提高检测分类器的性能, 在训练样本和测试样本中进行挖掘。

1.4.1 行人模型的训练

在训练行人模型时, 希望每一个特征块的位置和 LFVA 直方图相对稳定, 这样, 行人模型才更具一般性。为此, 采用分层聚类的方法。在行人包围盒中, 随机取若干个大小为 $h_f \times w_f$ 的矩形区域, 用式(4)计算每个矩形区域的 LFVA 直方图。对所有区域的 LFVA 直方图聚类, 得到初始分类。在其中, 进一步按照区域的左上角坐标聚类。在最终的聚类结果中, 每一类矩形区域有相似的 LFVA 直方图, 并且有相近的位置, 可看做是一个特征块。即, 第 i 类矩形区域对应第 i 个特征块 $F_i = (V_i, P_i)$, 其中 V_i 为第 i 类矩形区域的 LFVA 直方图的均值向量, P_i 为第 i 类矩形区域左上角点的中心。

1.4.2 检测分类器的训练

对检测分类器而言, 任何一个自然图像中都有大量的负样本 (不含有行人的检测窗口), 想获得好的检测效果, 要尽可能的降低对负样本的虚警率, 同时保证较高的检测率。为此, 在得到初始的检测分

类器后,还要在训练数据和测试数据中挖掘,以提高检测分类器的性能。

训练检测分类器分为3步,首先训练初始检测分类器。在训练数据集中,每一个含有行人的图像均标出了行人的包围盒,将其中的图像变换为 $h_w \times w_w$ 大小,与检测窗口的大小保持一致。然后,计算行人模型中每一个特征块 F_i 在其中出现的概率 $P(W_{x,y,l}|F_i)$,得到的数值组成一个向量,作为初始正样本。在每一个不含行人的图像中随机取若干个 $h_w \times w_w$ 大小的窗口。同样的,计算每个特征块在其中出现的概率 $P(W_{x,y,l}|F_i)$,得到的数值组成一个向量,作为初始负样本。初始正样本与初始负样本一起组成初始训练样本,训练一个AdBoost分类器,称为初始检测分类器 C_0 。

接下来,为了降低检测分类器的误报率,利用初始检测分类器在训练数据中挖掘困难负样本,并再次训练检测分类器。具体方法是对训练数据集中不含行人的图像进行顺序扫描,即每隔固定步长取一个大小为 $h_w \times w_w$ 的检测窗口,用检测分类器判断该其中是否有行人出现,如果结果错误,即 C_0 判断该窗口中存在行人,则将其加入困难负样本中。困难负样本与初始样本一起,训练检测分类器 C_1 。

最后,在测试数据集中挖掘,进一步提高检测分类器的检测率。用检测分类器 C_1 对测试数据进行测试,在判定为正(即检测分类器认为检测窗口中含有行人)的测试样本中,认为检测分类器输出数值较大的样本比较可信的。同样的,在判定为负(即检测分类器认为检测窗口中不含行人)的测试样本中检测分类器输出数值较小的样本也比较可信。称这些样本为可信测试样本,把可信测试样本与初始训练样本、困难负样本一起训练最终的检测分类器 C_2 。

经过挖掘后的检测分类器性能有很大的提升,详见2.2节。

1.5 检测

滑动窗口搜索是最常用的检测方法,亦采用此方法在图像中检测行人。对图像和尺度空间 Ω 中所有可能的窗口 $W_{x,y,l}$ 计算出现行人的概率 $P(W_{x,y,l}|\mathbf{M})$ 。在整个空间 Ω 中, $P(W_{x,y,l}|\mathbf{M})$ 的局部极大值点就是行人出现的位置。为找到此极值点,采用Mean Shift算法,其中核函数为高斯核函数。

2 实验结果

用INRIAPerson数据集对本文的方法进行测试,以检验其性能。下面分别介绍行人模型训练、检测分类器训练和行人检测的实验结果。

2.1 行人模型的训练

INRIAPerson数据集中,选取248幅行人图像用于训练行人模型。在每幅图像的行人包围盒中随机取40个块,按照1.4.1节中的方法聚类,共得到约1000个类,即有约1000个特征块。图4显示了块大小为 32×32 像素的部分聚类结果。其中,每一行为一个类的前5个样本。从图中可以看到,每一类包含了人体的一个特征部位,例如第1行代表了胸部的特征,第2行代表了脚部的特征,第4行代表了头部的特征,而最后一行则是腰部的特征。图5表示了图4中10个特征块的相对于行人包围盒的位置,其中,黑色的区域为行人包围盒,蓝色的矩形框是该类中每一个块的位置,红色的矩形框是它们的中心。从图5中可以看到,每一类特征块都收敛到一个相对集中的位置上。另外,由于人体的姿态和外观变化比较大,人体的同一部位可能会有不同的LFVA直方图,会产生不同的特征块。例如,图4中的第7行和第8行,都是表示人体的腿部的特征块,而由于腿部的着装不一样,产生了两个特征块。此种特性使得本文的模型能够识别不同外观的行人。

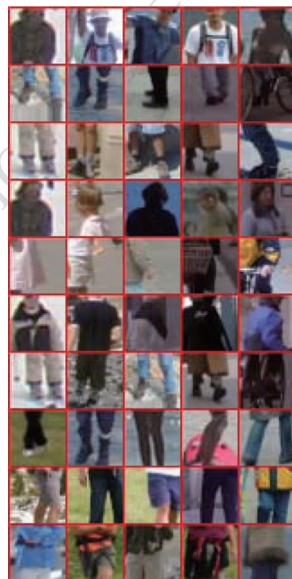
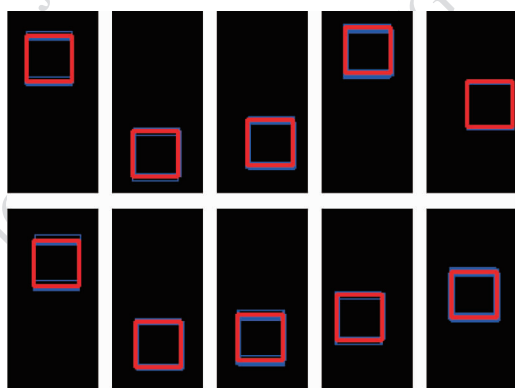


图4 部分特征块的聚类结果

Fig.4 Results of the feature blocks clustering



黑色的区域为行人包围盒,蓝色矩形为类中每一个样本的位置,红色的矩形为该类中样本的中心位置

图 5 特征块的位置特征

Fig. 5 Position of the feature blocks

2.2 检测分类器

利用 INRIAPerson 的训练数据集训练检测分类器。检测窗口大小设定为 64×128 像素。INRIAPerson 的训练数据集中有 1218 幅不含行人的图像,在每一幅图像中随机取 10 个检测窗口,据此计算初始负样本。训练得到初始检测分类器 C_0 后,在训练样本中用顺序扫描的方法得到约 1 万个困难负样本,与初始训练样本一起训练得到检测分类器 C_1 。在测试样本中,分别取得分排名前 0.5% 且被判定为正的样本,以及得分排名后 0.5% 且被判定为负的样本为可信样本,与初始训练样本、困难负样本一起训练得到检测分类器 C_2 。检测分类器 C_0 、 C_1 、 C_2 在 INRIAPerson 数据集中测试数据上的性能如图 6 所示(用于测试和训练的数据是分开的,没有交集)。从图中可以看到,检测分类器 C_1 和 C_2 比初始的检测分类器 C_0 的性能有了很大的提高,而 C_2 又比 C_1 的性能更好一些。

图 7 显示了不同参数配置对检测分类器性能的影响。其中,图 7(a)显示了计算 HoL 时子块直方图统计区间数量(bin)对检测分类器性能的影响。可以看到,随着 bin 数量的增大,HoL 提供了更多信息,检测分类器的性能不断提升。图 7(b)显示了 LFVA 两个圆周半径的不同配置对分类器性能的影响,总的来看, $r_1 = 0$ 时检测分类器的性能要略好于 $r_1 = 1$ 时的检测分类器。这是由于 r_1 、 r_2 的不同配置控制着 LFVA 特征对不同尺度的局部细节的表现能力,而在 INRIAPerson 数据集中图像大多清晰度较高,所以使用较小数值的 r_1 、 r_2 配置能突出小尺度的局部信息,有利于提高检测率。如果将 LFVA 用于

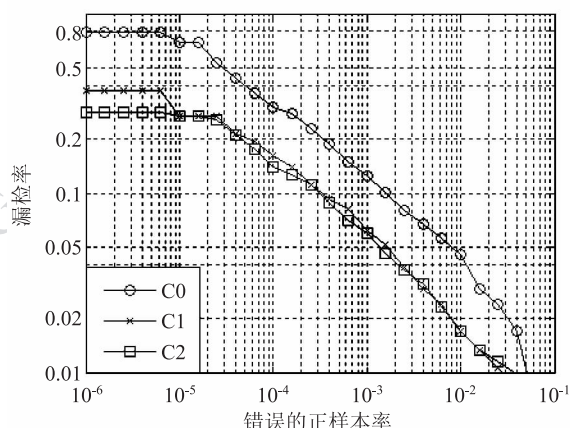


图 6 检测分类器在 INRIAPerson 数据集上的性能

Fig. 6 Performance of detection classifiers on INRIAPerson

分辨率较低的场景中进行行人检测,则最优 r_1 、 r_2 配置会有所不同。图 7(c)显示了式(7)中 b_0 、 b_1 的取值对检测分类器性能的影响, $b_0 = 0.6$ 、 $b_1 = 0.4$ 时检测分类器的性能是略好一些。这说明特征块的 HoL 与位置特征相比,略显重要一些。图 7(d)显示了特征块的大小对检测分类器性能的影响,特征块大小为 16×16 像素时检测分类器的性能要略好于其他的情况。这是由于在检测窗口为 64×128 像素时,行人的特征部位(例如头、胳膊、腿等)的宽度或高度接近 16 像素。故而,特征块大小为 16×16 像素时能较好的表现行人的特征部位。

图 8 显示了在最优参数配置下检测分类器 C_2 与文献[6]中检测分类器性能的对比,从图 8 中可以看到,在错误正样本率为 $10^{-4} \sim 10^{-3}$ 附近,两者性能差别不大;当错误正样本率小于 10^{-4} 时,HoG 方法要优于本文方法;在错误正样本率大于 10^{-3} 时本文检测分类器的性能则更好一些。

2.3 行人检测

图 9 显示了用本文方法在 INRIAPerson 数据集中检测行人的部分实验结果。图 10 显示了 HoG 方法^[6](在 OpenCV2.0 中实现)的部分实验结果。从图 10 中可以看到,HoG 在分辨率较低时容易漏检。例如,在 HoG 的检测结果中,图 10(e)的左上角区域有两个行人,由于距离较远,分辨率很低,没有被检测到。而本文方法由于利用了人眼的视觉注意原理,具有多尺度的特性,即使行人距离观察者很远、分辨率很低也能被检测到,见图 9(e)。从图 10 的结果中还能发现 HoG 方法对强的竖直边缘比较敏感,容易造成误检,例如图 10(b)——(d)(f)(j)(k)

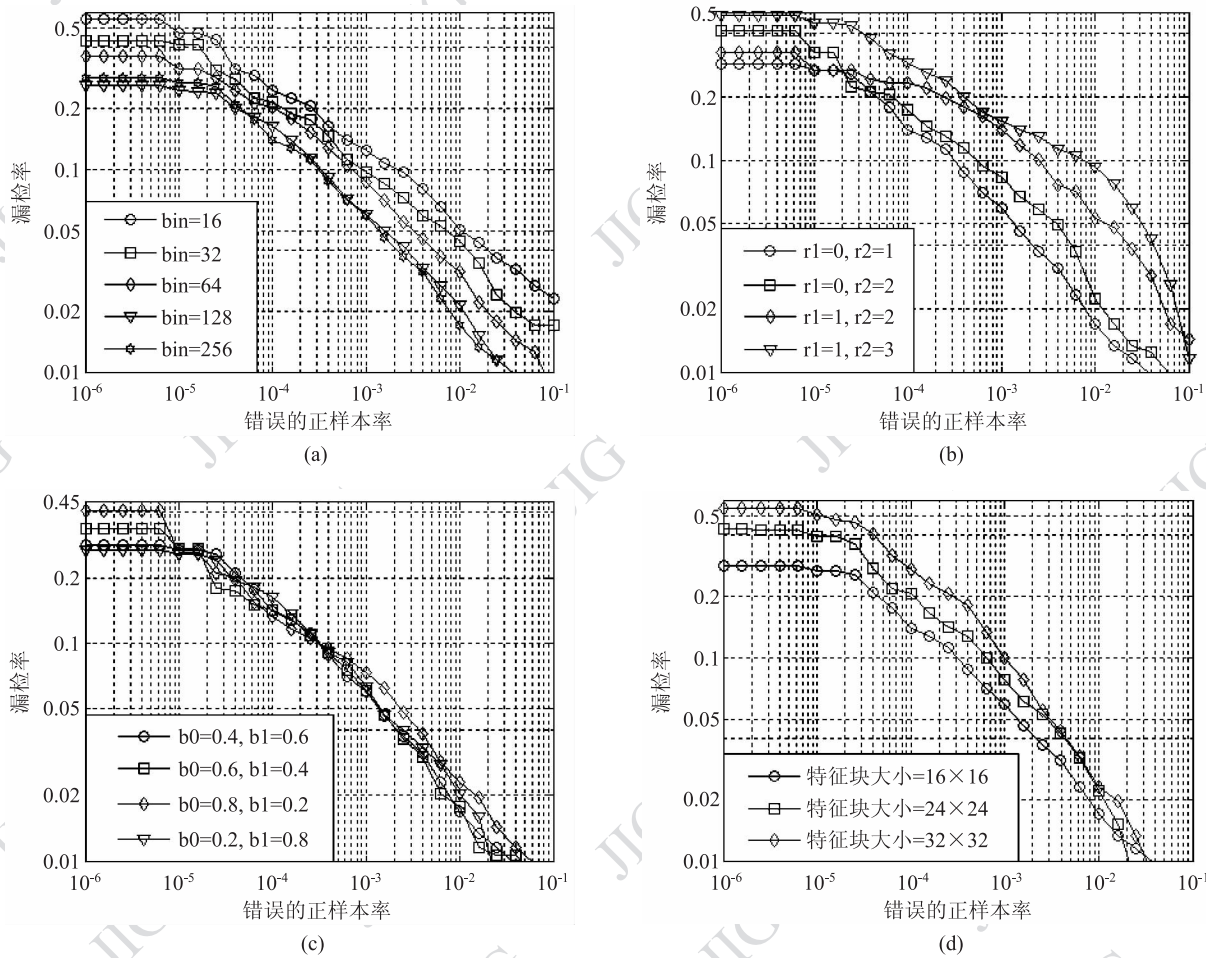


图7 不同参数对检测分类器的影响
Fig. 7 The effect of different parameters

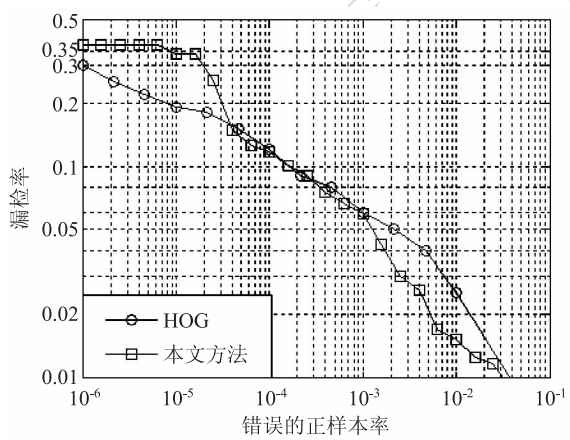


图8 本文方法与文献[6]方法的性能对比
Fig. 8 Comparison the performance of our method with reference [6]'s

中误检均发生在竖直边缘处。对比图 9, 本文方法的结果则不存在此类的问题。此外, 由于本文的方法采用了部分匹配的技术, 对不同姿态的行人以及有遮挡的情况检测也取得了比 HoG 方法好的结果。例如, 对比图 9(a) 和图 10(a), 玩耍的小孩的肢体运动幅度较大难以被 HoG 方法检测到, 而本文方法的结果则相对较好。对比图 9(e) 和图 10(e), 本文方法可以检测到被遮挡住腿部的女士, 而 HoG 方法则不行。

图 11 显示了本文方法与文献[3]中方法检测结果的对比。图 11(a) 是文献[3]方法在 PASCAL 数据集上对人体检测的最佳结果之一, 其中, 在行人较小、分辨率较低的位置出现了一些漏检。与之相比, 本文方法在此类位置的漏检较少。



图 9 本文方法的行人检测结果
Fig. 9 Some results of our method



图 10 文献[6]方法的部分检测结果
Fig. 10 Some results of reference [6]'s method



(a)文献 [3] 方法

(b)本文方法

图 11 文献[3]的结果与本文结果对比

Fig. 11 Comparing our result with reference[3]'s

3 结 论

目标检测是计算机视觉领域一个重要问题,而其中的行人检测问题又具有特殊的复杂性。与其他目标相比,行人由于肢体的运动产生了复杂的形状,由于衣着的不同产生了多变的外观,加之遮挡和视角变化,给行人模型及检测方法的设计带来了巨大困难。提出一种基于人眼视觉注意原理的局部特征表示方法,并将其用于行人检测。此局部特征利用了视觉注意原理,具有光照不变、旋转不变和多尺度的特性,能够处理低分辨率的情况。用基于特征块的模型表示行人,把行人看做特征块的集合,并利用部分匹配的方法检测行人。这种方法能够处理遮挡的情况以及因关节运动等带来的形状变化。与现有的方法相比,本文的方法具有对竖直边缘不敏感的优点,取得了较好的检测效果。在今后的研究中将重点考虑特征块之间的连接关系,以期能进一步提高行人检测的性能。

参考文献 (References)

[1] Duong T, Phung D, Bui H, et al. Efficient duration and hierarchical modeling for human activity recognition [J]. Artificial Intelligence, 2009, 173(7-8): 830-856.

[2] Liu H, Yu Z, Zha H, et al. Robust human tracking based on multi-cue integration and mean-shift[J]. Pattern Recognition

Letters, 2009, 30(9): 827-837.

[3] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object Detection with Discriminatively Trained Part Based Models[J]. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2009, 32(9): 1627-1645.

[4] Bai X, Wang X, Latecki L J, et al. Active skeleton for non-rigid object detection[C]//Proceedings of International Conference on Computer Vision. Kyoto, Japan: IEEE Computer Society, 2009: 575-582.

[5] Shotton J, Blake A, Cipolla R. Multi-scale categorical object recognition using contour fragments[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(7): 1270-1281.

[6] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA USA: IEEE Computer Society, 2005: 886-893

[7] Enzweiler M, Gavrila D M. Monocular pedestrian detection: survey and experiments [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2179-2195.

[8] Viola P, Jones M J, Snow D. Detecting pedestrians using patterns of motion and appearance[J]. International Journal of Computer Vision, 2005, 63(2): 153-161.

[9] Gavrila D M, Munder S. Multi-cue pedestrian detection and tracking from a moving vehicle [J]. International Journal of Computer Vision, 2007, 73(1): 41-59.

[10] Montabone S, Soto A. Human detection using a mobile platform and novel features derived from a visual saliency mechanism[J]. Image and Vision Computing, 2010, 28(3): 391-402.

[11] Palmer S E. Vision Science: Photons to Phenomenology [M]. Cambridge, MA USA: MIT Press, 1999: 28-35.

[12] Itti L. The ilab neuromorphic vision C++ toolkit; free tools for the next generation of vision algorithms[J]. The Neuromorphic Engineer, 2004, 1(1): 10.

[13] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

[14] Frintrop S, Hertzberg J. VOCUS. A Visual Attention System for Object Detection and Goal-directed Search [M]. Berlin: Springer, 2006: 55-86.

[15] Treisman A, Gelade G. A feature integration theory of attention [J]. Cognitive Psychology, 1980, 12(1): 97-136.

[16] Ojala T, Pietikainen M. Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.