



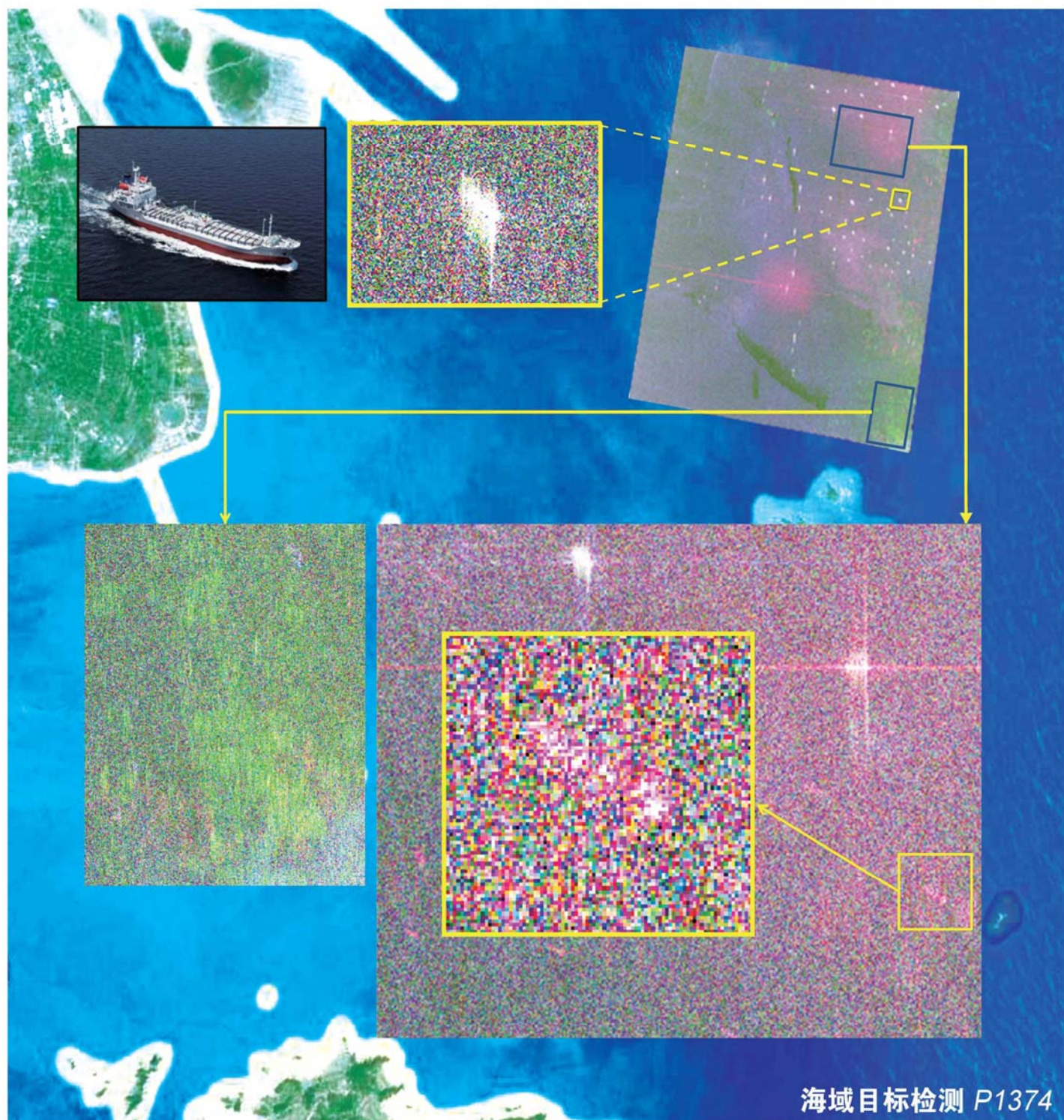
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

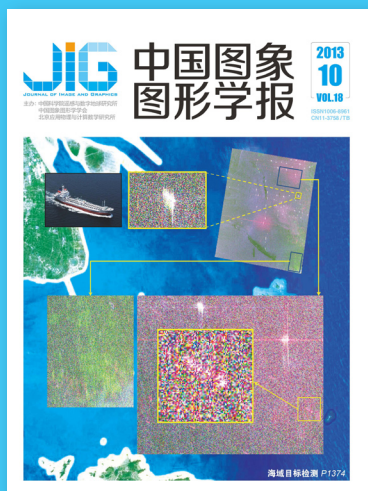
中国图象图形学报

2013
10
VOL.18

ISSN1006-8961
CN11-3758/TB



海域目标检测 P1374



第18卷第10期（总第210期）

2013年10月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

2013 all rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@irsa.ac.cn

电 话 010-64807995 010-82614429

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of
Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@irsa.ac.cn

Telephone 010-64807995 010-82614429

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

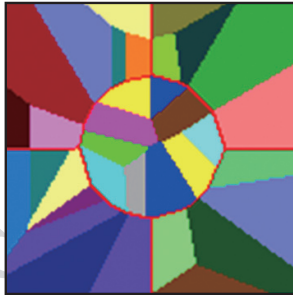
ISSN 1006-8961

CODEN ZTTXFZ

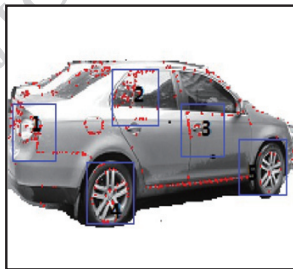
国外发行代号 M1406

国内邮发代号 82-831

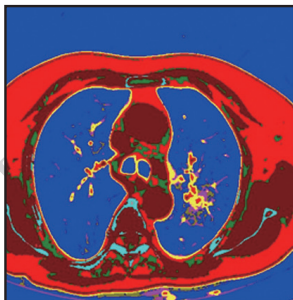
国内定价 50.00元



结合几何划分技术和最大期望值/最大边缘概率算法的彩色图像分割(第1270页)



联合生成与判别模型的目标检测与跟踪(第1293页)



聚类核值相似区特征点的医学影像分类(第1322页)

综述

手掌静脉图像识别技术综述

吴微, 苑玮琦 1215

视频检测烟雾的研究现状

罗胜, Jiang Yuzheng 1225

图像处理和编码

自适应分数阶微分的复合双边滤波算法

胡伏原, 姒绍辉, 张艳宁, 孙瑾秋 1237

曲波域高斯混合尺度模型的图像压缩重构

纪建, 李晓, 许双星 1247

结合图像信号显著性的自适应分块压缩采样

王瑞, 余宗鑫, 杜林峰, 万旺根 1255

四阶各向异性扩散方程在图像放大中的应用

李鹏, 邹杨, 姚正安 1261

图像分析和识别

结合几何划分技术和最大期望值/最大边缘概率算法的彩色图像分割

赵泉华, 李玉, 何晓军 1270

自适应加权完全局部二值模式的表情识别

胡敏, 许艳侠, 王晓华, 黄忠, 朱弘 1279

面向局部光照突变的时间和空间中心对称局部二值模式算子

卞紫涵, 房胜, 徐田帅 1285

联合生成与判别模型的目标检测与跟踪

刘倩, 侯建华, 牟海军, 赵巍, 笪邦友 1293

目标不同视角下观察信息的迁移和再利用

张索非, 吴海洋, 吴镇扬 1302

图像理解和计算机视觉

张量分析在视差图生成中的应用

赵戈, 蔺兰, 唐延东, 王耀南 1307

采用旋转匹配的二进制局部描述子

卢鸿波, 孙愿, 张志敏 1315

医学图像处理

聚类核值相似区特征点的医学影像分类

李博, 曹鹏, 栗伟, 赵大哲 1322

异常宫颈细胞核的自适应局部分割

张灵, 李静立, 陈思平, 汪天富, 江少锋, 刘少雄 1329

地理信息技术

Hilbert填充曲线与空间分布模式探测的点数据集空间划分方法

吴明光 1336

采用案例归纳推理进行道路网智能选取

郭敏, 钱海忠, 黄智深, 刘海龙, 王骁 1343

遥感图像处理

红外目标抗干扰跟踪算法

褚桐, 马惠敏 1354

MSTAR图像的矩特征分析与多阈值分割

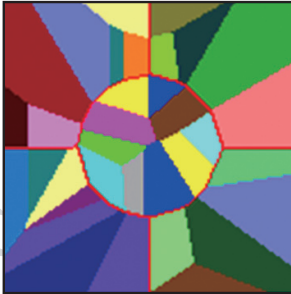
倪维平, 严卫东, 吴俊政, 郑刚, 芦颖 1364

改进Notch滤波的全极化SAR数据船舶检测方法

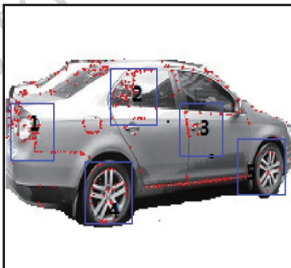
孙渊, 王超, 张红, 张波, 吴樊 1374

CONTENTS

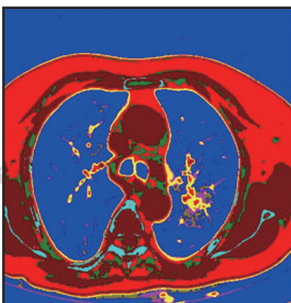
JOURNAL OF IMAGE AND GRAPHICS



Color image segmentation using geometry tessellation technique and EM/MPM algorithm(P1270)



Object detection and tracking combining generative and discriminative model(P1293)



Medical image classification based on cluster univalue segment assimilation nucleus feature point(P1322)

Review

- A survey of palm-vein image recognition
Wu Wei, Yuan Weiqi 1215
- State-of-art of video based smoke detection algorithms
Luo Sheng, Jiang Yuzheng 1225

Image Processing and Coding

- Multi-bilateral filtering algorithm based on adaptive fractional order differentiation
Hu Fuyuan, Si Shaohui, Zhang Yanning, Sun Jinqiu 1237
- Image reconstruction algorithm based on the curvelet gaussian scale mixture model
Ji Jian, Li Xiao, Xu Shuangxing 1247
- Saliency-based adaptive block compressive sampling for image signals
Wang Rui, Yu Zongxin, Du Linfeng, Wan Wanggen 1255
- Fourth-order anisotropic diffusion equations for image zooming
Li Peng, Zou Yang, Yao Zheng'an 1261

Image Analysis and Recognition

- Color image segmentation using geometry tessellation technique and EM/MPM algorithm
Zhao Quanhua, Li Yu, He Xiaojun 1270
- Facial expression recognition based on AWCLBP
Hu Min, Xu Yanxia, Wang Xiaohua, Huang Zhong, Zhu Hong 1279
- TSCS-LBP operator oriented to the local illumination mutation
Bian Zihan, Fang Sheng, Xu Tianshuai 1285
- Object detection and tracking combining generative and discriminative model
Liu Qian, Hou Jianhua, Mou Haijun, Zhao Wei, Da Bangyou 1293
- Transfer and reusing of object view information
Zhang Suofei, Wu Haiyang, Wu Zhenyang 1302

Image Understanding and Computer Vision

- Disparity map generation with tensor analysis
Zhao Ge, Lin Lan, Tang Yandong, Wang Yaonan 1307
- Binary local descriptor based on rotative matching
Lu Hongbo, Sun Yuan, Zhang Zhimin 1315

Medical Image Processing

- Medical image classification based on cluster univalue segment assimilation nucleus feature points
Li Bo, Cao Peng, Li Wei, Zhao Dazhe 1322
- Segmentation of abnormal cervical nuclei using an adaptive and local approach
Zhang Ling, Li Jingli, Chen Siping, Wang Tianfu, Jiang Shaofeng, Liu Shaoxiong 1329

Geoinformatics

- Hilbert space-filling curve and spatial pattern detection-based spatial partitioning approach to point geospatial data
Wu Mingguang 1336
- Intelligent road network selection method based on cases inductive reasoning
Guo Min, Qian Haizhong, Huang Zhishen, Liu Hailong, Wang Xiao 1343

Remote Sensing Image Processing

- Study of infrared anti-jamming tracking algorithm
Chu Tong, Ma Huimin 1354
- Moment feature analysis of MSTAR image and multiple thresholds based segmentation
Ni Weiping, Yan Weidong, Wu Junzheng, Zheng Gang, Lu Ying 1364
- Polarimetric SAR ship detection using improved notch filter
Sun Yuan, Wang Chao, Zhang hong, Zhang bo, Wu Fan 1374

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2013)10-1293-09

论文引用格式: 刘倩, 侯建华, 牟海军, 赵巍, 笄邦友. 联合生成与判别模型的目标检测与跟踪[J]. 中国图象图形学报, 2013, 18(10): 1293-1301. [DOI: 10.11834/jig.20131010]

联合生成与判别模型的目标检测与跟踪

刘倩^{1,3}, 侯建华^{1,3}, 牟海军^{1,3}, 赵巍², 笄邦友^{1,3}

1. 中南民族大学电子信息工程学院, 武汉 430074; 2. 华中农业大学食品科技学院, 武汉 430070;
3. 中南民族大学智能无线通信湖北省重点实验室, 武汉 430074

摘要: 提出一种新的基于生成-判别模型的目标检测与跟踪方法。利用 DAISY 特征描述子所具有的对光照、形变、视角、尺度的不变性以及计算高效的特点, 提取目标稳定的特征点并加以表达, 形成生成模型; 采用霍夫森林分类器作为判别模型, 用以训练目标图像块。在后续视频序列中利用目标的检测结果和判别码本的相似性测量对模型进行更新, 构建一个动态自适应的判别码本。实验结果表明这种将快速有效的 DAISY 描述子和识别率高、鲁棒性强的霍夫森林分类器相结合的算法, 跟踪精度高、实时性较好, 具有目标局部防遮挡能力和不同分辨率下的识别能力。

关键词: 目标检测与跟踪; 生成模型; 判别模型; DAISY 特征描述子; 霍夫森林

Object detection and tracking combining generative and discriminative model

Liu Qian^{1,3}, Hou Jianhua^{1,3}, Mou Haijun^{1,3}, Zhao Wei², Da Bangyou^{1,3}

1. College of Electronic Information Engineering, South-Central University for Nationalities, Wuhan 430074, China;
2. College of Food Science & Technology of Huazhong Agricultural University, Wuhan 430070, China;
3. Hubei Key Laboratory of Intelligent Wireless Communications, South-Central University for Nationalities Wuhan 430074, China

Abstract: A new object detection and tracking method is proposed based on generative and discriminative models. First, stable features of object are extracted and represented by the DAISY feature descriptor, which has computational efficiency and is invariant to illumination, deformation, viewpoint, and scale. In this way, the object generative model is constructed. Second, the Hough forest classifier is adopted as discriminative model and the input patches of object are trained. Moreover, the discriminative codebook is updated by computing the similarity measurement between the detection results of the following video sequence and the codebook, and makes the codebook dynamic and adaptive. Experiment results show that the proposed algorithm, combining DAISY feature descriptor and Hough forest, has satisfactory tracking precision and good real-time performance. Additionally, it works well under the condition of partial occlusions and different image resolutions.

Key words: object detection and tracking; generative model; discriminative model; DAISY feature descriptor; Hough forest

收稿日期: 2012-12-08; 修回日期: 2013-03-19

基金项目: 国家自然科学基金项目(61141010); 武汉市科技供需对接计划项目(201051824575); 湖北省自然科学基金项目(2010CDB02001, 2012FFA113); 中南民族大学智能无线通信湖北省重点实验室开放课题(IWC2012005)

第一作者简介: 刘倩(1987—), 女, 现为中南民族大学通信与信息系统专业硕士研究生, 主要研究方向为视频监控和机器视觉。E-mail: liuqian-198768@163.com

通讯作者: 侯建华, E-mail: zil@scuec.edu.cn

0 引言

视频序列中目标的准确检测与跟踪是一个颇具挑战性的任务,其中制约跟踪效果的最大因素莫过于目标的外表变化,包括姿势和形状等内因变化,以及光照、移位、相机视角及遮挡等外因变化。因此,寻找合适的特征,建立稳定的目标模型,准确地把目标与背景分离是视频目标检测与跟踪中一个至关重要的问题。

目标检测与跟踪方法一般可以分为基于生成式模型和基于判别式模型两大类^[1-2]。基于生成式模型的方法通过构建一个自适应模型,在指定的特征空间中对目标区域进行描述,即提取目标样本的特征,对特征的统计分布建模,再利用模型进行预测或匹配。因此该方法的重点在于能否寻找到一些好的特征来表述目标,建立一个尽可能涵盖目标各种状态(不同视角,不同尺度,有形变等)的模型。从本质上说该方法是基于运动目标信息的建模。例如经典的卡尔曼滤波^[3]就是根据运动方程对下一时刻的目标状态做出预测与更新;LK跟踪算法^[4]则是根据运动前后图像灰度保持不变这个基本假设,利用像素的运动信息构造光流约束方程,寻找错误率最小的匹配;文献[5-7]是利用具有尺度、旋转、仿射不变性的特征描述子表示目标表面的特征集;文献[8]则是利用直方图信息来建描述目标模型。

基于判别式模型的方法是将目标检测视为一个目标和背景的二分类问题,旨在寻找不同类别之间的最优分类面和判决函数,使得目标与背景的差异性最大。文献[9]利用支持向量机(SVM)作为分类器,把目标(正样本)从背景(负样本)中分离出来,这是判别式模型的典型代表。判别式方法的关键是分类器设计,早期的方法是在跟踪前训练好目标分类器,这使其应用受限于已知目标的跟踪。近年来,自适应判别式检测跟踪器^[10-11]设计引起了广泛的关注,这种方法的特点是在跟踪过程中构建分类器(作为检测器辅助跟踪),具有更好的鲁棒性能。

一般说来,基于生成式模型的方法中由于特征描述和匹配过程的计算量大,难以满足实时性要求。同时,该方法过多强调目标建模而忽略了背景因素的影响,容易导致对场景变化缺乏适应性。而基于判别分类器的方法大多需要丰富的训练样本,即训练样本必须包含不同视角、状态的目标表面,这在实

际应用中是很难实现的。

为了能够兼顾两者的优势,研究者尝试把生成式和判别式结合起来^[1,12-14]。文献[1]提出利用序列贝叶斯学习把生成式和判别式结合起来的方法,用序列混合高斯模型(SGMM)建立生成模型,利用快速相关向量机(FRVM)建立判别模型,该方法能够很好地解决复杂场景的目标检测与跟踪问题。Kalal等人在文献[12]中提出了一种Tracking-Learning-Detection算法,其中跟踪(Tracking)是利用中值流(median-flow)法估计目标物的运动模型,检测(Detection)是用随机森林分类器建立判别模型将目标从背景中分割出来,通过在线P-N学习^[15],将两者联系起来,对判别模型实时更新。该方法的稳定性和鲁棒性都很好,但在目标旋转、分辨率发生变化和局部遮挡时效果不佳。

提出一种新的基于生成-判别模型的方法。首先将目标图像划分成若干图像块(patch),用高斯差分(DoG)法检测出目标图像中稳定的特征点,利用计算效率高、具有不变性的DAISY^[16]特征描述子来描述这些图像块的特征点,然后将图像块和其对应的类别标记作为霍夫森林(Hough forest)分类器的训练数据,通过结构性的训练和学习,生成霍夫森林相应的结构参数,把目标图像块和背景图像块分开。霍夫森林的各叶节点就形成了判别式的码本集。为了提高目标模型的自适应性,提出一种新的模型更新方法,即在后续视频序列中,利用新标记的图像块与模型码本中图像块的相似性测量对模型进行更新。实验结果证明,将快速有效的DAISY描述子和识别率高、鲁棒性强的霍夫森林分类器相结合,在处理局部遮挡、消失目标重现检测和低分辨率图像等场合具有显著的效果,同时新的模型自适应更新的方法使算法对复杂场景下的目标检测与跟踪有很大的优势。

1 生成模型

为了使目标模型能够尽可能地涵盖目标的表面变化,提出首先利用多尺度空间的高斯差分^[17]检测图像中的特征点,再用DAISY特征描述子对图像进行描述,建立一个具有旋转不变性、视角不变性、尺度不变性,对光照和局部遮挡有一定稳定性的模型。

1.1 特征提取

提取稳定的特征是目标模型描述的首要任务,

要求提取的特征点应含有显著的结构信息, 这些点可以没有实际的直观视觉意义, 但却在某种角度、某个尺度上含有丰富稳定的信息。采用 DoG 方法提取图像中的特征点, 这种方法是用不同尺度(标准差)的高斯函数对图像进行平滑, 通过对平滑后的图像进行比较, 将差别大的像素作为稳定的特征点。DoG 的实现步骤如下:

1) 构建高斯金字塔^[5]。首先采用不同尺度因子的高斯核对图像进行卷积以得到图像的不同尺度空间, 将这一组图像作为金字塔图像的第1层。

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y) \quad (1)$$

式中, $L(x, y, k\sigma)$ 为不同尺度下的图像尺度空间表示, k 是同一层中相邻两个图像的尺度因子的比例系数, $I(x, y)$ 表示空间位置 (x, y) 处的像素值, 高斯核为

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$$

对第1层中第1个图像以2倍像素距离下采样得到金字塔第2层中的第1个图像, 对该图像采用不同尺度因子的高斯核进行卷积, 获得第2层塔的一组图像。这样依次类推, 从而获得了金字塔图像的每一层中的一组图像。

2) 采用高斯差分的方法检测局部位置的极值点, 即同一层中两个尺度相邻的图像相减, 即

$$D(x, y, \sigma) =$$

$$(G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (2)$$

3) 对 DoG 尺度空间每个点与相邻尺度和相邻位置的点逐个进行比较, 得到的局部极值点就认为是在该尺度下的一个特征点。

4) 最后, 通过拟合3维二次函数以精确确定关键点的位置和尺度, 再去除边界和低亮度区域处不好的特征点。

1.2 特征点描述

特征点描述的方法有很多, 最常用的就是尺度不变特征变换(SIFT)^[17], 这种描述子首先确定关键点的位置和关键点所处的尺度, 然后使用关键点邻域梯度的主方向作为该点的方向特征, 以实现算子对尺度和方向的无关性。SIFT 的缺点是计算量大、耗时, 为此提出了一些改进算法, 比如 PCA-SIFT^[18] 和 SURF^[19] 算法, 这些算法在一定程度上提高了效率, 但同时也使描述子的稳定性下降。

2008 年 Engin 等人提出了一种高效的特征描述方法——DAISY^[16] 特征描述子, 它不仅继承了 SIFT 的不变性优势, 而且相对于改进的 PCA-SIFT 和

SURF 算法来说运算效率高, 鲁棒性强, 对遮挡不敏感, 对图像分辨率要求不高。

DAISY 描述子是利用特征点及其邻域信息来描述的。具体结构如图1所示。它使用一组不同方向不同核宽(标准差)的高斯核与原图像的卷积来实现图像梯度的加权过程。

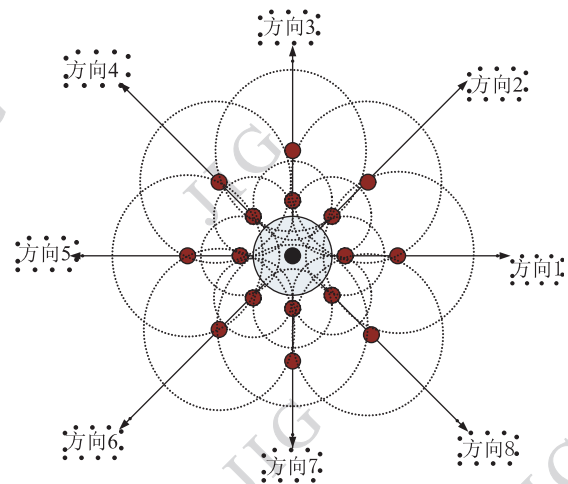


图1 DAISY 描述子结构

Fig. 1 The structure of DAISY descriptor

图1中, 黑色圆点代表中心特征点, 每个圆圈代表特征点周围的一个区域, 红色圆心代表邻域采样点的位置, 圆的半径与高斯核的标准差成比例关系, 高斯核的大小则表征了对梯度进行加权因子的大小, 在离中心特征点近的地方用小的高斯核模糊, 距离远的地方用大的高斯核模糊, 通过改变高斯核的核宽可以有效地去除较远位置梯度模的影响。同时圆心也代表卷积后的方向图的中心位置。所以, 相比较 SIFT/SURF 来说, DAISY 采用的是圆环采样网格, 用不同核宽的高斯核和图像梯度的卷积来实现对梯度的加权, 由于高斯核具有各向同性, 所以当要计算不同方向的描述子时, 不需要重新进行卷积的计算。

DAISY 描述子的生成过程如下:

1) 给定一幅已提取出特征点的图像, 首先计算8个方向的梯度方向图

$$G_o = \left(\frac{\partial I}{\partial o} \right)^+ \quad (3)$$

式中, o 表示梯度的某一个方向, G_o 就表示在 o 方向的梯度值。

2) 通过核宽为 W 的高斯核 G_w 与梯度方向图的卷积运算来对梯度加权, 得到 o 方向的卷积方向图

$$G_o^W = G_W * G_o \tag{4}$$

从图 1 中可以看出每个方向上都被核宽不同的高斯核卷积多次,得到不同区域的卷积地图。高斯卷积同时也去除了噪声。

实际上,用大的高斯核作卷积可以从一些小的连续高斯核卷积得到,例如已知核宽为 W_1 的高斯核,求核宽为 W_2 的高斯核与梯度方向图的卷积,则

$$G_o^{W_2} = G_{W_2} * G_o = G_{\sqrt{W_2^2 - W_1^2}} * G_{W_1} * G_o = G_{\sqrt{W_2^2 - W_1^2}} * G_o^{W_1} \tag{5}$$

正是由于高斯卷积的这种迭代性质使得 DAISY 算子非常高效,可以用于实时性要求高的应用场合。

3) 对于图像中某一特征点 (μ_0, ν_0) , 其统计直方图可以用 H 维的卷积方向图矢量得到

$$\mathbf{h}_w(\mu_0, \nu_0) = [G_1^W(\mu_0, \nu_0), \dots, G_H^W(\mu_0, \nu_0)]^T \tag{6}$$

式中, $G_1^W, \dots, G_H^W$ 是不同方向的方向图的高斯核, H 是方向数(例如图 1 中取的是 8 个方向)。然后再把 $\mathbf{h}_w(\mu_0, \nu_0)$ 归一化,记为 $\tilde{\mathbf{h}}_w(\mu_0, \nu_0)$ 。

4) 最后,DAISY 描述子是由中心点和它的邻域采样点的卷积矢量来形成。设 Q 为以中心特征点为圆心的同心圆的个数, $i=1, 2, \dots, Q; j=1, 2, \dots, H$, 则

$$\mathbf{D}(\mu_0, \nu_0) = [\tilde{\mathbf{h}}_{w_1}^T(l_1(\mu_0, \nu_0, R_1)), \dots, \tilde{\mathbf{h}}_{w_Q}^T(l_Q(\mu_0, \nu_0, R_Q))]^T$$

$$\begin{aligned} & \tilde{\mathbf{h}}_{w_2}^T(l_1(\mu_0, \nu_0, R_2)), \dots, \tilde{\mathbf{h}}_{w_2}^T(l_T(\mu_0, \nu_0, R_2)), \\ & \vdots \\ & \tilde{\mathbf{h}}_{w_Q}^T(l_1(\mu_0, \nu_0, R_Q)), \dots, \tilde{\mathbf{h}}_{w_Q}^T(l_T(\mu_0, \nu_0, R_Q)) \end{aligned} \tag{7}$$

$\mathbf{D}(\mu_0, \nu_0)$ 是 $Q \times H$ 的分块矩阵,矩阵中的基元为 $\tilde{\mathbf{h}}_w^T(l_j(\mu_0, \nu_0, R_i))$, 表示在 j 方向上与中心特征点 (μ_0, ν_0) 距离为 R_i (第 i 层同心环的半径)的邻域采样点的卷积矢量。

每层的高斯核的核宽 σ_i 和半径 R_i 为

$$\begin{aligned} \sigma_i &= \frac{R(i+1)}{2Q} \\ R_i &= \frac{R(i+1)}{Q} \end{aligned} \tag{8}$$

式中, R 是最外层的圆环的半径。

由于 DAISY 描述子特殊的圆环形结构和高斯卷积本身的优势,使得 DAISY 计算效率高,可应用性强,表 1 是对 3 幅图像的所有像素点分别用 SIFT、PCA-SIFT、SURF 和 DAISY 描述的计算时间。从表 1 可以看出,相对于 SIFT,DAISY 方法的计算时间有近 100 倍的提高,相对于 PCA-SIFT、SURF 均有不同程度的提高,这对于动态视频中的目标检测是很有意义的。

表 1 SIFT 和 DAISY 的计算时间
Table 1 The computing time of SIFT and DAISY

图像大小/像素	DAISY	SIFT	PCA-SIFT	SURF
800 × 600	3.8	252	243	83
1 024 × 768	6.5	432	417	149
1 280 × 960	9.8	651	630	217

1.3 建立生成模型

为了提高目标的防遮挡能力,对目标采用多个局部的图像块来表示,分别对这些图像块的特征点进行 DAISY 描述。

图 2 中目标(汽车)有 N 个图像块,利用 DoG 的方法对每个图像块提取特征。随机从中选取 M 个特征点来表示图像块,这样第 n 个图像块就可以用所提取的特征点表示为

$$\mathbf{I}_n = \{k_{n1}, k_{n2}, \dots, k_{nm}, \dots, k_{nM}\}$$

用 DAISY 描述子 I_n^{DY} 对这些特征点进行描述

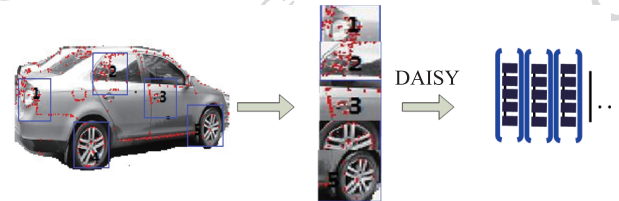


图 2 生成模型

Fig. 2 Generative model

$$I_n^{\text{DY}} \leftarrow \text{DAISY}(k_{n1}, k_{n2}, \dots, k_{nm}, \dots, k_{nM})$$

那么目标的表面集模型 Model 可表示为

$$\{I_1^{\text{DY}}, I_2^{\text{DY}}, \dots, I_n^{\text{DY}}, \dots, I_N^{\text{DY}}\}$$

2 判别模型

通过将目标的特征用 DAISY 方法进行描述,建立了稳定的目标生成模型,将生成模型作为霍夫森林分类器的输入,通过霍夫森林的学习获得目标(前景)和背景的判别函数,在对后续的视频序列进行检测,并更新目标模型,使目标和背景尽可能地分离。

2.1 HF 的训练

2009 年 Gall 提出了霍夫森林^[20]的概念,霍夫森林是在随机森林(RF)^[21]的框架基础上,把霍夫变换^[22-23]引入以实现图像特征空间到霍夫参数空间的映射。霍夫变换损失的信息少,对遮挡具有一定免疫力,所以和 RF 相比,HF 除了具有随机森林分类器高效,无过拟合、鲁棒性强的优势外^[24-25],对目标的局部遮挡也有很强的鲁棒性。

和随机森林一样,霍夫森林也是由一系列决策树组合而成,每颗决策树的叶节点存储了霍夫变换所需的投票信息。霍夫森林中的决策树以有监督的方式训练,每棵树的训练过程兼顾了随机化和最优化。森林中用于训练的样本是以 Bagging 的方式随机采样的,从训练样本中随机抽取 k 个子样本集,建立 k 棵决策树;最后,根据这 k 棵决策树的分类对每个测试样本进行投票决定其最终类别。

如图 3 所示,每一棵决策树由一个根节点、众多分支节点和叶节点组成,每颗决策树从它对应的根节点开始,根据不同的特征属性进行分裂,到达下层分支节点的其中之一,再分裂,直到满足节点分裂的条件或达到设定的最大树深,这样形成一个叶节点;到达同一个叶节点的是同一类别的局部图像块,相应的信息被存储在对应的叶节点上。

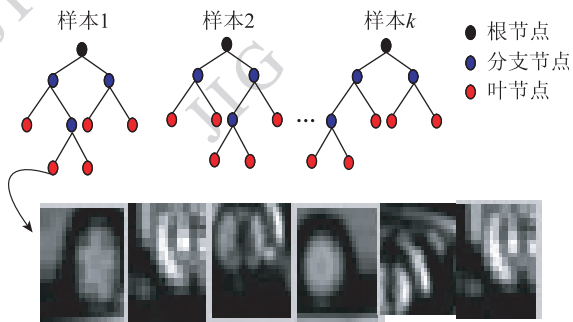


图 3 随机决策树

Fig. 3 Random trees

在本文中,霍夫森林的训练样本是从视频帧中采样的局部图像块,从目标中(边界框内部的部分)采样的分配类别 $c_n = 1$,从背景中(边界框外)采样分配类别 $c_n = -1$ 。在构建决策树时,随机地从训练样本中选取 k 个训练集,构建 k 棵树,每个图像块包含的信息表示为 $\{\rho_n = (I_n, c_n, d_n)\}$, I_n 是由 DAISY 描述的表面特征集, c_n 是图像块的类别,为了保留各个图像块之间的空间相关性,文中引入了一个参量 d_n ,它表示该图像块中心到目标跟踪框中心的距离。

在训练开始时,将训练的样本集加入到每棵决策树相应的根节点,自上而下逐分支分裂,当图像块到达某个叶节点 L 时,该图像块的信息就存储到该叶节点,叶节点保存了两种 p_L 和 D_n 信息,其中 p_L 表示到达叶节点 L 的图像块中属于目标类别的比例, D_n 指目标类别所有可能的偏移的集合, $D_n = \{d_1, d_2, \dots, d_n\}$ 。这样每棵树的叶节点可看做一个有识别力的码本。如何利用这两个信息,进行霍夫变换,是需要解决的问题。

在对视频序列中的车辆或行人进行跟踪时,目标在相邻帧之间的运动是不会急速变化的,鉴于此,在选取后续视频帧中选取测试样本时,在边界框两倍的范围内搜索,提高正样本的比例。为方便起见,把测试样本取为二值测试样本,其定义是通过比较一对像素点的值的大小来形成,即 $t(I) \rightarrow \{0, 1\}$

$$t_{i,j,p,q} = \begin{cases} 0 & I(i,j) < I(p,q) \\ 1 & \text{其他} \end{cases} \quad (9)$$

式中, $I(i,j)$ 表示在位置 (i,j) 的像素值,将这些二值测试样本分配给决策树的每一个非叶节点,根据分类原则送入下一层的节点继续分裂直到到达某个叶节点 L ,则对应的测试样本就拥有该叶节点的信息。

2.2 Hough 投票

对于随机森林来说,最终的决策结果是求取每棵决策树结果的平均,而霍夫森林是将广义的霍夫变换引入到最终的投票中。

在目标检测时,需要利用图像块的投票确定目标边界框的中心位置。设一个霍夫空间 H ,霍夫空间 H 中的基元由关于中心位置 x_c 和尺度 s 的假设集 $\{h(x_c, s)\}$ 组成。到达某叶节点 L 的图像块,就认为该图像块拥有叶节点的所有属性,即该图像块属于目标类别的概率是 p_L ,可能出现的位置离目标中心的偏移是 D_n 。它以概率 p_L 和位置偏移量 D_n 对霍夫空间中相应的元素进行投票。虽然每个投票元素图像块不提供目标的精确位置信息,但是将来

自不同投票元素的像素累加成一个霍夫图像,通过检测霍夫图像中的峰值,就可以确定目标中心,其中峰值的高度表征了目标可能的位置和尺度的置信度。

给定初始的目标边界框 B_x , 把图像块的信息按照空间位置定义为 $\rho(y) = (I(y), c(y), d(y))$, 即中心位于 y , 偏移量为 $d(y) = y - x_c$ 。需要注意, 若 y 不在 B_x 内部, 则 $I(y)$ 肯定不属于目标, 那么对于定位目标是没有意义的, 所以只考虑 y 在 B_x 内部, 即属于目标的 $I(y)$, 其叶节点 $L(y)$ 对于霍夫空间的投票。这样利用已知目标样本的局部特征 $I(y)$ 和相应的叶节点信息 $L(y)$ 确定目标出现的位置和尺度, $h(x_c, s)$ 即寻找一个条件概率 $p(h(x_c, s) | L(y))$ 。具体的理论推导过程如下:

考虑只使用 $c(y) = 1$ 的样本对目标定位, 则

$$\begin{aligned} p(h(x_c, s) | L(y)) &= \\ p((h(x_c, s), c(y) = 1) | L(y)) &= \\ p(d | ((c(y) = 1, L(y)) \cdot p(c(y) = 1 | L(y))) & \quad (10) \end{aligned}$$

式中, $p(c(y) = 1 | L(y))$ 是目标的类别概率分布, 另一项 $p(d | (c(y) = 1, L(y)))$ 是类的空间概率分布。这两项就是判别模型中存储的信息。其中第 1 项以偏移量 d 为参量利用 Parzen 窗估计的方法, 对图像块的类别空间分布进行估计, 第 2 项就是训练过程中到达叶节点的图像块中属于目标的比例 p_L 。参数 S_u 是归一化的尺度系数。

$$\begin{aligned} p(h(x_c, s) | L(y)) &= \\ p((h(x_c, s), c(y) = 1) | L(y)) &= \\ \frac{1}{|D_n|} \left[\frac{1}{S_u} \sum_{d \in D_n} \exp\left(-\frac{\|(y - x_c) - d\|^2}{2\sigma^2}\right) \right] p_L & \quad (11) \end{aligned}$$

设森林中决策树的个数为 T , 统计所有决策树的分类结果得到

$$\begin{aligned} p(h(x_c, s) | I(y); \{T_i\}_1^T) &= \\ \frac{1}{T} \sum_{i=1}^T p(h(x_c, s) | L(y); T_i) & \quad (12) \end{aligned}$$

最后, 将来自不同图像块的投票结果累加之后得到霍夫图像

$$V(x_c, s) = \sum_{y \in B_x} p(h(x_c, s) | I(y); \{T_i\}_1^T) \quad (13)$$

计算出 $V(x_c, s)$ 的最大值和相对应的 $h(x_c, s)$, 那么 x_c 和 s 就是检测的目标的中心和尺度, $V(x_c, s)$ 就是该假设 $h(x_c, s)$ 的置信度。

图 4 是霍夫投票的实验结果, 图 4(a) 第 1 行是经过预处理特征可视化的图像块, 第 2 行是相应的图像块对于目标边界框中心的投票, 即式(11)对应的实验结果。图 4(b) 是由所有图像块通过相应叶节点的类别信息构建的目标, 黄色圆点聚集的位置就是目标中心的范围, 黄色直线表示图像块相对于它所投票的目标中心的偏移量; 图 4(c) 第 1 行是由目标中心定位出的实际目标, 第 2 行表示霍夫空间中的目标中心, 在霍夫图像中亮度越大, 表示目标中心存在的可能性就越大。

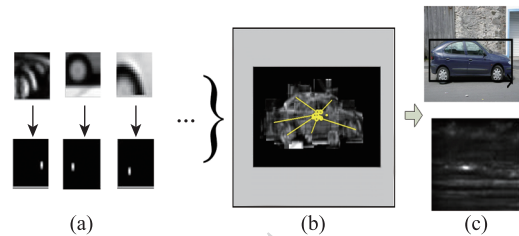


图 4 各图像块对于边界框中心的投票

Fig. 4 The votes of every patch to bounding box

2.3 模型更新

霍夫森林采用的是离线训练方式, 在训练完成后, 参与投票的判别码本信息 p_L 和 D_n 是固定的, 而在个体目标跟踪的场合, 目标的变化很大, 单从初始帧获取训练数据来建立目标模型是不够完善的, 所以提出在检测过程中实时地对目标模型进行更新。

从式(11)可以看出, 判别码的类别概率 p_L 相当于对目标的投票的加权因子, 在训练完成后, 每个叶节点的 p_L 是固定的。提出一种更新策略, 在后续的检测过程中对 p_L 进行更新, 方法如下:

首先, 当从后续视频帧中提取的未知标记的图像块遍历随机森林中的所有决策树, 找到图像块到达的使 p_L 最大的叶节点, 即满足 $p_{\max} = \max(p_L)$, 将该叶节点对应的码本记为 $C_{\max} = \{b_1, b_2, \dots, b_m\}$, m 是一个叶节点所包含的图像块的总数。对于新到达的某个图像块(记为 b'), 分别计算它与 $C_{\max}$ 中每个码本元素的相似度, 两个图像块 b_i, b_j 的相似度定义为

$$S(i, j) = 0.5(NCC(b_i, b_j) + 1) \quad (14)$$

式中, $NCC()$ 为归一化协方差函数。

这里给 $S(i, j)$ 设定一个经验阈值 $\theta = 0.6$, 统计码本中和 b' 相似度大于 0.6 的码本基元的总数 N , 若 $N > \frac{m}{2}$, 则把新 b' 加入该码本中, 为了保持码本数

的平衡,随机丢弃原码本库中的一个图像块。这样就得到新的码本 C ,相应地算出该叶节点的 p'_L 和 D'_L 。最终,通过这种更新机制,保证了目标模型跟最近的时间内的目标状态相似度最大。更新机制的伪代码如下:

输入 由霍夫森林训练得到的 N 个码本投票信息 p_{L_i} 、 D_{L_i} 。其中 $i=1,2,\dots,N$ 。

输入 未标记的 b' 。

输出 更新后的码本 C' 相应投票信息 p'_L 、 D'_L 。

$m \leftarrow 0$ //初始化参数

1) 未标记的图像块 b' 遍历霍夫森林,到达森林中每棵树的某个叶节点 $\{L_1, L_2, \dots, L_i, \dots, L_N\}$, 叶节点对应携带的投票信息为 $\{p_{L_1}, p_{L_2}, \dots, p_{L_i}, \dots, p_{L_N}\}, \{D_{L_1}, D_{L_2}, \dots, D_{L_i}, \dots, D_{L_N}\}$

2) 计算投票元素 p_{L_i} 的最大值

$p_{\max} \leftarrow \max(p_{L_i})$

$C_{\max} \leftarrow C(\max(p_{\max}))$

3) 统计所有满足下列条件的 m

For $C_{\max}$ 中所有的图像块

do while $0.5(NCC(b', b_m) + 1) \geq 0.6$

$m \leftarrow m + 1$

4) If $m \geq \frac{N}{2}$ then

$C' \leftarrow C_{\max} \cup b'$ 。

随机舍弃一个 $C_{\max}$ 中的图像块。

返回更新后的 C' 和投票信息 p'_L 、 D'_L 。

end if

end for

图5是针对模板更新机制的验证实验,实验中采用的数据是 AVSS 2007 中的 AVSS AB Medium 视频序列的前 900 帧。图5(a)是目标跟踪效果的直

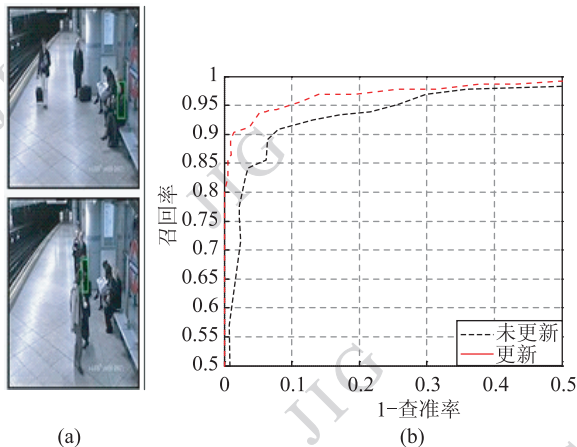


图5 模板更新的对比实验

Fig. 5 Comparison between no-updated and updated model

观表示,图5(b)是分别根据采用更新机制前后的实验结果所绘制的召回率和1-查准率性能曲线。可以看出,和固定模型的方法相比,模型更新后(红色实线)的精度有明显提高。

3 实验步骤

本文方法主要用于对个体目标进行检测与跟踪,在首帧通过鼠标响应选定跟踪区域,对该区域进行仿射变换得到训练的正样本集,从背景中选取负样本来训练霍夫森林。把后续视频序列作为测试数据,对目标进行检测,并利用 Condensation 算法^[26]对目标进行预测跟踪,在跟踪时刻 t ,相应的状态变量由目标的中心、尺度及边界框的速度形成 $S_t: (x_t, s_t, v)$ 。

如图6所示,视频1是利用摄像头获取的视频序列,视频中扇子被多角度的旋转,或近或远的偏离摄像头。图6(a)是鼠标选取女孩的脸作为跟踪目标,这里训练集是通过目标图像进行仿射变换得到的一系列图像集。图6(b)(c)是将扇面进行了90°的旋转,从不同视角观察目标,可以看出,在视角发生变化时,仍能定位出目标;图6(d)是将目标逐渐接近摄像头,尺度系数为2.0,在分辨率发生变化时目标也没有丢失。

图7视频是从不同方位对目标进行不同程度的遮挡,从实验结果可以看出,算法也能够准确地跟踪

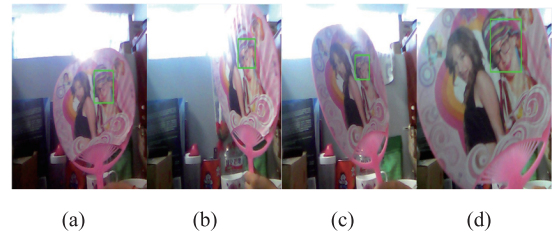


图6 不同视角不同尺度下的跟踪结果

Fig. 6 Tracking in different views and scales

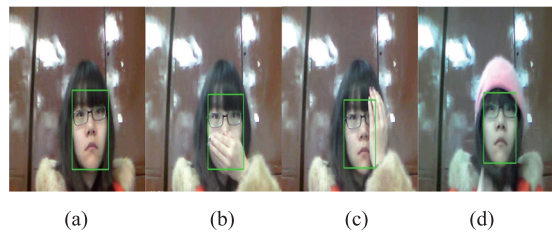


图7 局部遮挡情况下的跟踪结果

Fig. 7 Tracking with partial occlusion

到目标,这是由于算法中 DAISY 算子是一个局部特征,同时 Hough 变换本身有一定的防局部遮挡能力。

图 8 视频是不同光线下对目标的测试结果,目标逐渐由光线较暗的室内走到光线强的室外的过程中,目标的模型也逐渐被更新,保证目标总是和最临近的状态相似度最大。

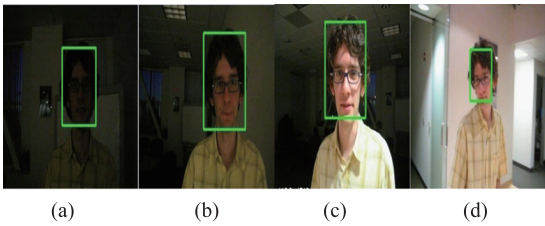


图 8 不同光线下的跟踪结果

Fig. 8 Tracking with changes in illumination

为了测试算法在复杂场景下的性能,截取了一

段某车载镜头获得的视频(图 9),这段视频共 659 帧,背景是实时变化的,存在光照,周围车辆的干扰,车速时快时慢,车辆转弯,目标车辆左右行驶等变化。实验结果表明,在这种真实场景下,算法依然能够准确地跟踪目标(图 9),具有很强的鲁棒性。

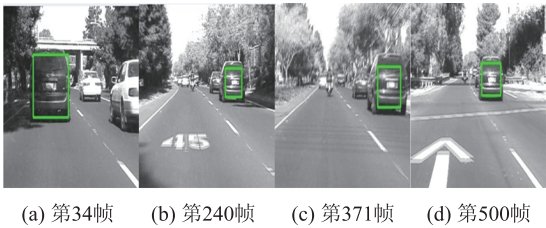


图 9 复杂场景下的跟踪结果

Fig. 9 Tracking in cluttered background

表 2 是 4 组视频在变化的场景(不同尺度、分辨率、光线和有局部遮挡)的情况下的实验分析,并分别统计了每个视频的识别率和帧速。

表 2 算法性能分析

Table 2 Analysis of performance

视频	帧数/帧	不同尺度	不同分辨率	局部遮挡	不同光线	识别率/帧	帧速/(帧/s)
图 6	536	✓	✓	—	✓	536	21
图 7	842	✓	—	✓	—	838	20
图 8	770	✓	✓	—	✓	739	16
图 9	659	✓	✓	✓	✓	651	18

注:“✓”表示能够处理该变化;“—”表示对应视频中本身不存在该场景变化。

4 结 论

利用具有尺度不变性的 DoG 特征点检测 + DAISY 特征描述来构建目标模型,使得模型具有局部特征的防遮挡能力、视角不变性、旋转不变性,对图像分辨率不敏感的稳定特性。此外在随机森林分类器的基础上,引入了 Hough 变换,通过 Hough 投票对目标进行精确定位,并提出一个高效准确的模型更新方法,形成了一个综合生成模型和判别模型的目标检测跟踪算法。可以广泛的应用于行人或人脸检测,交通监测等诸多领域。

参考文献(References)

[1] Lei Y, Ding X Q, Wang S J. Visual tracker using sequential bayesian learning: discrinnative, generative, and hybrid [J]. IEEE Transactions on System Man And Cybernetics, 2008,

38(6):1578-1591.
[2] Dinh T B, Medioni G. Cotraining framework of generative and discriminative trackers with partial occlusion handling [C]//Proceedings of IEEE Workshop on Applications of Computer Vision. Los Angeles, USA: IEEE, 2011: 642-649.
[3] Curwen R W, Amini A A, Duncan J S. Tracking vascular motion in x-ray image sequences with kalman snakes [C]//Proceedings of Computers in Cardiology. Los Angeles, USA: IEEE, 1994: 109-112.
[4] Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision [C]//Proceedings of International Joint Conference on Artificial Intelligence. San Francisco, USA: IEEE, 1981, 18(5): 674-679.
[5] Lowe D G. Object recognition from local scale-invariant features [C]// Proceedings of IEEE International Conference on Computer Vision . Corfu, Greece: IEEE, 1999, 2: 1150-1157.
[6] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, USA: IEEE, 2003, 27(10): 257-263.

- [7] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509-522.
- [8] Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006, 1: 798-805.
- [9] Avidan S. Support vector tracking[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2004, 26(8): 1064-1072.
- [10] Grabner H, Bischof H. Online boosting and vision[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006, 1: 260-267.
- [11] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009: 983-990.
- [12] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [13] Lin R S, Ross D, Lim J, Yang M H. Adaptive discriminative generative model and its applications[C]//Proceedings of Neural Information Processing Systems. Vancouver, Canada: IEEE, 2004: 801-808.
- [14] Yang M, Wu Y. Tracking nonstationary appearances and dynamic feature selection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005, 2: 1059-1066.
- [15] Kalal Z, Matas J, Mikolajczyk K. P-N learning: Bootstrapping binary classifiers by structural constraints[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010: 49-56.
- [16] Tola E, Lepetit V, Fua P. A fast local descriptor for dense matching[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008: 1-8.
- [17] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [18] Ke Y, Sukthankar R, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE, 2004: 506-513.
- [19] Bay H, Tuytelaars T, Gool L V. Surf: speeded up robust features[C]//Proceedings of European Conference on Computer Vision. Graz, Austria: IEEE, 2006: 404-417.
- [20] Gall J, Lempitsky V. Class-specific Hough forests for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009: 1022-1029.
- [21] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [22] Ballard D H. Generalizing the Hough transform to detect arbitrary shapes[J]. Pattern Recognition, 1981, 13(2): 111-122.
- [23] Leibe B, Leonardis A, Schiele B. Robust object detection with interleaved categorization and segmentation[J]. International Journal of Computer Vision, 2008, 77(1): 259-289.
- [24] Lepetit V, Lagger P, Fua P. Randomized trees for realtime keypoint recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005: 775-781.
- [25] Bosch A, Zisserman A, Munoz X. Image classification using random forests and ferns[C]//Proceedings of IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007: 1-8.
- [26] Yao J M. Study on particle filter based visual tracking method[D]. Changchun: Changchun Institute of Optics Fine Mechanics and Physics, 2004: 50-81. [姚剑敏. 粒子滤波跟踪方法研究[D]. 长春: 长春光学精密机械与物理研究所, 2004: 50-81.]