

基于对象性测度估计和霍夫森林的目标检测方法

胡梦婕, 魏振忠, 张广军

(北京航空航天大学 精密光机电一体化技术教育部重点实验室, 北京 100191)

摘要: 实现高效准确的目标检测算法在视频监控、自动导航等诸多领域有着至关重要的作用。针对现有目标检测算法速度不高且鲁棒性差的缺点, 提出了一种基于对象性测度估计和霍夫森林的快速目标检测方法。首先, 基于自下而上的视觉注意机制, 采用对象性测度估计, 提取图像中的物体候选集; 然后, 在由物体候选集确定的感兴趣区域内进行霍夫森林目标检测, 确定目标中心位置; 最后, 结合目标中心所在的对象性测度估计候选框的尺度信息, 确定目标大小。实验结果表明, 该方法在提高霍夫森林目标检测算法检测准确度的同时, 大大提升了检测速率。

关键词: 目标检测; 霍夫森林; 对象性测度估计

中图分类号: TP391.4 文献标志码: A 文章编号: 1007-2276(2015)06-1936-06

Object detection method based on objectness estimation and Hough forest

Hu Mengjie, Wei Zhenzhong, Zhang Guangjun

(Key Laboratory of Precision Opto-mechatronics Technology, Ministry of Education, Beihang University, Beijing 100191, China)

Abstract: Realizing effective and efficient object detection plays an important role in computer vision and has many practical applications including video surveillance and auto navigation. In order to improve the speed and accuracy of the existing detection methods, a simple yet effective object detection method coupled objectness estimation with Hough forest was proposed. Firstly, objectness estimation was utilized to generalize a set of object proposals based on bottom up visual attention mechanism of human vision system. Secondly, Hough forest object was adopted to localize the center of the object in the region of interest which was confirmed by object proposals put forward in the last step. Thirdly, the scale of the object proposal where the center was located was employed to determine the size of the object. A set of experiments demonstrate the effectiveness and efficiency of the proposed method.

Key words: object detection; Hough forest; objectness estimation

收稿日期: 2014-10-13; 修订日期: 2014-11-17

基金项目: 国家自然科学基金(61340054); 北京市自然科学基金(3142012); 国家重大科学仪器设备开发专项(2012YQ140032)

作者简介: 胡梦婕(1988-), 女, 博士生, 主要研究方向为计算机视觉与模式识别等。Email: humengjie881101@gmail.com

导师简介: 张广军(1965-), 男, 中国工程院院士, 主要研究方向为计算机视觉与模式识别等。Email: gjzhang@buaa.edu.cn

0 引言

据统计,人们从外部世界获取的信息 80% 以上来自于视觉^[1],计算机视觉的目的就是使用摄像机和电脑代替人眼进行目标识别、跟踪和测量,并进行场景和行为理解。

目标检测作为高层视觉任务的基础,是计算机视觉中至关重要的一个研究问题,被广泛地应用于视频监控、自动导航、身份验证等多个领域。目标检测主要分为建模阶段和检测阶段。建模阶段通常分为两个步骤:(1)建立待检测目标的表征模型;(2)利用数据库里的样本图片采用机器学习方法获得表征模型参数。检测阶段,对于新入图片采用滑动窗的方法进行检测,使用非极大抑制获得最优目标位置。

然而,感知心理学和神经学的研究者都指出人类在识别物体之前会先感知到物体。传统的滑动窗目标检测方法不仅违背了这种人类视觉系统对图像的理解机制,而且非常耗时。显著性检测和对象性测度估计均为立足于先感知后识别的视觉注意机制。显著性检测试图分割出场景图像中最显著的物体,但是当一幅图像中存在多个物体并且物体在图像中不占主要成分时,显著性检测的效果并不理想。对象性测度估计(Objectness estimation)是用来衡量一个图像窗口内是否包含隶属于任何类别的物体的参数。现在国际上先进的对象性测度估计算法均通过寻找提高物体判别性的复杂特征来进行物体检测^[2-3],因此速度限制成为了实时应用的瓶颈问题。近期,Cheng^[4]提出了一种简单有效的特征描述——二值化赋范梯度(BING)用来进行对象性测度估计,算法运行速度可达到 300 fps,使得实时应用中将视觉注意机制作为目标检测的前端成为了可能。

广义霍夫变换为目标检测中的一类主流方法,由检测到的物体独立部分为物体中心投票,检测假设对应着由所有部分投票获得的霍夫图像的极值点。基于广义霍夫变换和随机森林,Gall^[5-6]提出了霍夫森林算法,在随机森林的框架下,通过有监督学习将图像块的外观和霍夫票选建立了直接的映射关系,由于霍夫森林的高效性和容错性,它被广泛应用于目标检测、跟踪及行为识别。然而,在霍夫森林建立的过程中,没有考虑尺度不变性问题,因此用来训

练的图像块采集自尺度归一化的图像数据集;在检测过程中,为了应对尺度变化问题,使用一系列尺度系数对测试图像进行尺度调整,并在每个尺度下单独计算相应的霍夫图像。此后,一系列霍夫图像被放置于三维尺度空间中,并用三维高斯滤波获得空间极值,寻找到对应尺度,进而求解目标尺度。但是此方法本质上是在进行尺度尝试,使得尺度空间过于离散化,而且计算量与尺度个数成线性关系,计算效率低下。因此,估计每个图像块的尺度问题成为了基于霍夫票选机制的算法的固有弱点。安萌等^[7]采用霍夫森林进行空间有形目标的检测与识别的过程中也不可避免地遇到了目标尺度问题,使用尺度尝试来判断目标大小。

为了解决这个问题,该文将对象性测度估计和霍夫森林相结合,解决了霍夫森林目标检测的尺度问题,在提升检测速率的同时,大大降低了目标检测的误检率。

1 算法概述

霍夫森林算法在物体检测时存在尺度适应的固有缺陷:首先,在检测过程中需要将随机采集图像块的尺度转换到霍夫森林训练时的尺度上,然而大部分情况下,在检测过程中计算机无法获得物体的尺度信息,因此只能将尺度离散化,通过尺度尝试挑选出效果好的尺度因子,从而确定目标的大小;其次,在霍夫票选出物体中心后,由于对应尺度信息不精确,因此无法准确定位出物体所在区域。这样的尺度尝试方法既耗时又不准确。

为了解决上述问题,文中提出了一种基于对象性测度估计和霍夫森林的目标检测算法,算法流程图如图 1 所示。总体来看,该算法主要由对象性测度检测部分和霍夫森林检测部分组成。

在对象性测度检测部分,通过样本库训练出描述所有物体共性的检测器。由于赋范梯度特征的引入,该检测器能够在输入图像中以 300 fps 的速度检测出可能是完整物体所在的区域,实验验证,对象性测度检测提出的前 1 000 个候选窗口具有 96.2% 的物体检测率。

在霍夫森林检测部分,通过样本对霍夫森林进行离线训练,获得霍夫森林目标检测器。在检测过程

中,通过先验知识(例如物体宽高比)在对象性测度检测部分提出的候选区域中筛选出感兴趣的物体区域,并根据霍夫森林训练样本的尺度对区域进行尺度变换,最后通过霍夫森林检测器对感兴趣区域进行检测,如果对应霍夫图像的极值大于一定阈值,则存在目标,并输出目标的区域。

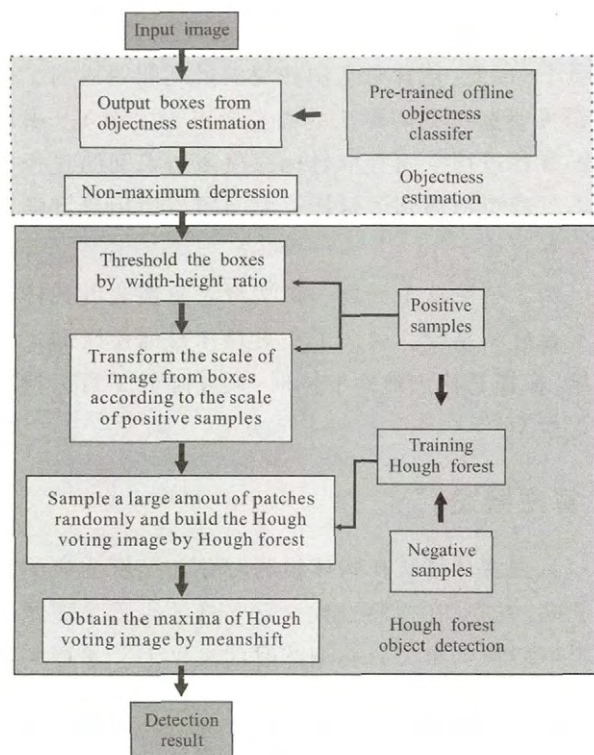


图 1 算法流程图

Fig. 1 Processing diagram of our proposed algorithm

2 对象性测度估计

Cheng 发现对于所有具有闭合轮廓的物体,当把物体对应的图像窗口归一化到一个固定的尺度,物体在归一化梯度空间有很强的相关性。为了寻找图像中的物体,Cheng 提出了一个 64 维的赋范梯度(Normed Gradient, NG)特征,用一系列不同大小(尺度和宽高比)的量化窗口对图像扫描,每一个窗口用一个线性模型 $\omega \in \mathbb{R}^{64}$ 来计算窗口得分 S_i :

$$S_i = \langle \omega, g_i \rangle \quad (1)$$

$$l = (i, x, y) \quad (2)$$

式中: S_i, g_i, l, i 和 (x, y) 分别表示一个窗口滤波后的分数, NG 特征, 位置, 大小和中心。对每一个大小为

i 的窗口,使用非极大抑制筛选出一小部分物体候选集。同时,考虑到有一些大小的窗口(例如 10×500)包含物体的可能性比另一些大小的框口(例如 100×100)低,因此,对象性测度估计分数最终可定义为:

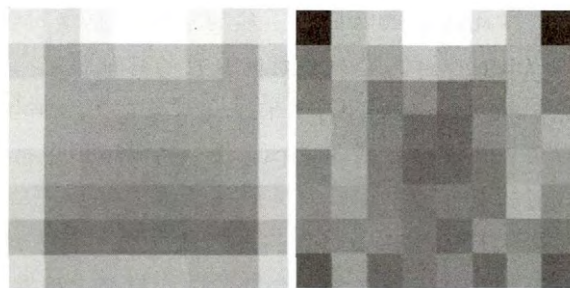
$$O_i = v_i \cdot S_i + t_i \quad (3)$$

式中: $v_i, t_i \in \mathbb{R}$ 分别表示对于每一个量化尺度 i 学习系数和偏置项。

2.1 赋范梯度和对象性测度

物体是具有完整闭合边界和中心的独立事物。当将真实世界中的物体在图像上的对应窗口调整到一个较小的固定大小(例如 8×8),在这样抽象概括的视图下,闭合边界会呈现出很小的变化,对应梯度图像的范数(如强度)显示出了很强的相关性,如图 2(a)展示了一个物体样本的 NG 特征,图中可以清楚地看出明显的近似圆形的模式。通过这个原理,笔者可以高效地预测物体的存在性。首先将输入图像调整到几个量化的大小上并且计算每一幅重设大小后的图像的 NG。这些调整大小后的 NG 图中的区域被定义为其对应窗口的 64 维 NG 特征。

NG 特征对平移、尺度和宽高比变化具有一定的不变性,并且 NG 特征的密集紧凑性使其具有相当高的计算效率,具有实时应用的潜力。但是, NG 特征对物体的判别性较弱,即能够划分物体区域,但无法区分物体类别,因此文中在物体显著性检测划分出物体区域后,通过霍夫森林实现某个物体类别的判断,并降低了物体性检测过程中生成的虚警。

图 2 NG 特征及学习得到的 ω Fig. 2 Normed gradient and the training result of ω

2.2 基于 NG 特征的对象性测度学习

通过采用两层级联 SVM 分类器来学习图像窗口的物质性测度,具体步骤如下:

- (1) 使用线性 SVM^[8]分类器学习公式(1)中的

单个模型 ω 参数,从图片测试库标准前景窗口中采集得到的 NG 特征作为正样本,从背景部分随机采集的 NG 特征作为负样本,输入到线性 SVM 进行训练;

(2) 为了使用线性 SVM 分类器学习公式(3)中的 v_i 和 t_i ,对于训练图片,在大小为 i 的前提下评价公式(1),使用候选集作为训练样例,滤波分数为一维特征,并且使用训练图像注释来验证他们的标签。

2.3 关于 ω 的讨论

如图 2(b)所示,学习得到的线性模型 ω 看上去与生物学界认可的灵长类架构的多尺寸中心环绕模式假设非常相似,在 ω 的边界权重较大表明模型倾向于寻找一个将物体和背景分开的分界线。此外,受益于模型二值化估计的优势,为了加速特征提取和测试过程,Cheng 提出了二值化赋范梯度,在此不予赘述,在参考文献[4]中可以找到具体叙述。

3 霍夫森林的建立与目标检测

3.1 霍夫森林的建立

与随机森林一样,霍夫森林是许多决策树的集合,但霍夫森林在随机森林学习框架的基础上,结合了位置偏执信息,每一颗决策树由一系列图像 $\{P_i=(I_i, c_i, d_i)\}$ 块构成,其中 I_i 表示图像块的特征描述符, c_i 是图像块类别标签, d_i 表示图像块距离目标中心的位移偏置向量。训练图像块是从带有前景目标标识的样本数据库中采集的,将从前景区域采集到的图像块的类别标签分配为 $c_i=1$,而从背景区域中采集到的图像块的类别标签 $c_i=0$ 。同时,对于背景图像块,记录图像块距离目标中心的位移偏置向量对于目标中心的定位没有意义,因此,设置位移偏置向量 $d_i=0$ 。

构建决策树时,每颗决策树由样本数据集的一个随机子集训练,以递归的方式从根部开始迭代,每一个节点都会接收到一系列图像块。对于非叶节点,按照指定的分裂法则寻找最优分割函数将到达此节点的图像块分成两个部分分别流向两个对应子节点。以此类推,决策树继续迭代生长。

选取节点分割函数的准则是尽量降低两个子节点数据的不确定性。到达决策树中每一个节点的信息 $A=\{P_i=(I_i, c_i, d_i)\}$ 包含图像描述符,标签和位移偏

置信息,因此霍夫森林定义了两个不确定度评价方式:标签类不确定度和偏置不确定度。标签类不确定度 $U_1(A)$ 用来衡量该节点类别标签的不纯度:

$$U_1(A)=|A| \cdot \text{Entropy}(\{c_i\}) \quad (4)$$

其中一系列均值为 c 的二值标签 $\{c_i\}$ 的信息熵为:

$$\text{Entropy}(\{c_i\})=-c \cdot \log c-(1-c) \cdot \log (1-c) \quad (5)$$

偏置不确定度 $U_2(A)$ 用来衡量位移偏置信息的不纯度:

$$U_2(A)=\sum_{i: c_i=1}\left(d_i-d_A\right)^2 \quad (6)$$

式中: d_A 为位移偏置的平均值。节点分裂公式随机选取一个不确定度评价方式使其最小化,对节点进行分割。当树的深度达到最大值或者节点图像块的数量足够少时,声明此节点为叶节点,并存储票选信息 C_L (来自目标的图像块的比例)和位移偏置向量集合 D_L 。

3.2 基于霍夫森林的目标检测

假设在训练和检测时目标框的大小均为 $W \times H$,在检测时,只需要确定矩形框的中心坐标即可。从待测图片上提取图像块,对于每一块中心为 y 的图像块 $P(y)=((I(y), c(y), d(y)))$,将对应的图像描述符 $I(y)$ 按照决策树每个节点的分割法则依次向下迭代,直到抵达叶节点,得到的分类结果表示为 $\{b_1, b_2, \dots, b_T\}$,其中 T 为决策树的总数, b_t 为图像块在随机森林第 t 个决策树上隶属的叶节点。最后,在霍夫图像中用分类结果进行投票,将所有 $I(y)$ 在随机森林的每棵树上的隶属节点 b_t 所对应的 C_L/D_L 累加到霍夫图像中位置为 $\{y-d|d \in D_L\}$ 的像素值上。累加每一个像素的票选结果,并对累加图像进行高斯滤波,得到霍夫图像,最后,通过均值迁移等方法获得霍夫图像的极值,即检测到的目标中心位置。

3.3 尺度问题

由于霍夫森林检测过程中固有的尺度问题,文中将对象性测度估计输出的候选集进行尺度滤波,生成小部分感兴趣区域,然后将每个感兴趣区域的尺度调整到霍夫森林训练样本的尺度上,进而输入到霍夫森林中进行霍夫票选。对每个霍夫子图像进行高斯滤波,若滤波后的图像极值大于设置的阈值,则认为对应的子候选框包含待测物体,候选框的大小即为物体大小;若滤波后的图像无明显极值,则认为此候选框不包含待测目标。

该方法采用对象性测度估计的结果对霍夫森林的尺度进行评测,不仅能够增强霍夫森林目标检测的鲁棒性,而且大大提升了霍夫森林的检测速度。

4 实验结果与分析

为了测试该方法的有效性,文中进行了一系列的实验验证,实验所用图像数据库来自 VOC 2007^[9],包含 20 个物体类别共计 9 963 张图片。软件采用 Visual Studio2010 开发,在 Windows7 环境下运行。

实验中,对象性测度估计训练器采用 4 952 幅图像进行训练,测试时每幅图像输出 1 000 个候选框,具有 96.2% 的物体覆盖度;霍夫森林检测器采用 35 幅人体图作为正样本,50 幅背景图作为负样本进行训练。霍夫森林训练中,对样本的尺度进行归一化,宽度统一为 70 pixel,训练层数为 10 层,采集图像块大小为 16×16 pixel,霍夫图像中物体存在性的阈值设置为 80。

检测时,从图像库中随机选择图片进行测试,部分测试结果如图 3、图 4、图 5 所示。图 3(a)、图 4(a)、

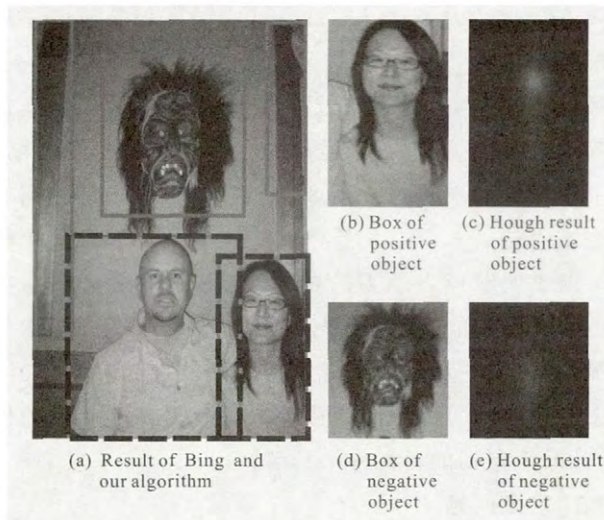


图 3 该文方法在 VOC2007 图像集上的测试结果。图(a)的矩形框为对象性测度估计得到的部分物体候选集,虚线矩形框为最终检测结果,图(b)为目标框,图(c)为目标框对应的霍夫图像,图(d)为非目标框,图(e)为非目标框对应的霍夫图像

Fig. 3 Object detection result of our algorithm, the images under test are from benchmark of VOC2007. (a) Detection boxes from Bing, and dash line boxes indicate the result of our algorithm; (b) Box of positive object; (c) Hough result of positive object; (d) Box of negative object; (e) Hough result of negative object

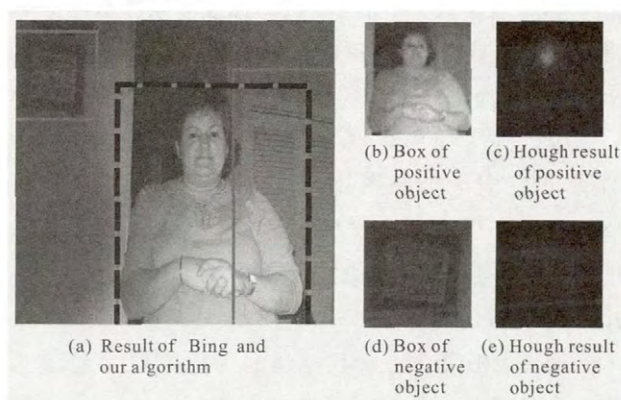


图 4 该文方法在 VOC2007 图像集上的测试结果。

子图像表示的意义与图 3 中的子图像相同

Fig. 4 Object detection result of our algorithm with the images from VOC2007. Sub-image meaning are same of Fig.3

图 5(a)中,矩形框表示对象性测度估计输出的部分候选集(图中只显示出了目标框和部分背景框),对候选集进行尺度滤波后,将得到的感兴趣区域的尺度调整到霍夫森林训练时的尺度(宽度为 70 个 pixel),如图 3(b)3(d)、图 4(b)4(d)、图 5(b)5(d)所示。然后通过霍夫森林对这些感兴趣区域进行物体检测,经过高斯滤波后的霍夫图像如图 3(c)3(e)、图 4(c) 4(e)、图 5(c)5(e)所示。从图中可以看出,包含人体的感兴趣区域计算得到的霍夫图像具有明显的中心,且中心的极值大于阈值;而其他感兴趣区域(不包含人体)对应的霍夫图像没有明显的中心,且最大值小

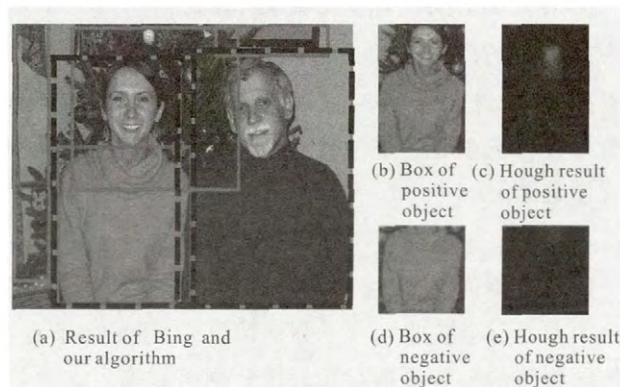


图 5 该文方法在 VOC2007 图像集上的测试结果。

子图像表示的意义与图 3 中的子图像相同

Fig. 5 Object detection result of our algorithm with the images from VOC2007. Sub-image meaning are same of Fig.3

于阈值。进而将含有中心极值的霍夫图像对应的候选框定为目标框,如图3(a)、图4(a)、图5(a)中虚线矩形框所示。

实验结果表明,该文提出的方法通过采用对象性测度估计较好地解决了霍夫森林算法的尺度问题,使得传统的基于广义霍夫变换的方法在检测过程中不再需要进行人为尺度尝试,大大提高了计算速度(传统方法对于分辨率375×500图像进行5个尺度检测的处理时间达到18.01 s,而此文的方法从对象性测度获得候选框,检测时间为2.4 s)。同时,由于只在对象性测度估计输出的候选集上计算霍夫图像,滤除了部分背景干扰,从而提高了目标检测的准确率。

5 结 论

该文提出了一种基于对象性测度估计和霍夫森林的目标检测方法。通过将视觉注意机制引入到目标检测的前端,在提升目标检测速度的同时,大大降低了检测误检率。该方法可以通过加强对象性测度估计算法来进一步提高检测准确率,文中采用了NG特征来描述所有物体的共性,简单有效,但是为了达到较高的物体覆盖率,需要采集1 000个左右候选框,通过尺度滤波形成400个左右的候选集,后续可通过寻找快速但鉴别力更强的特征来进行物体描述,降低目标候选集的数量,提升算法性能。

参考文献:

- [1] Xia Dong. Research on object detection, tracking and recognition in intelligence video surveillance [D]. Changsha: National University of Defense Technology, 2012: 1-2. (in Chinese)
 - [2] Alexe Bogdan, Deselaers Thomas, Ferrari Vittorio. Measuring the objectness of image windows [J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2012, 34(11): 2189-2202.
 - [3] Wang Xiaoyu, Yang Ming, Zhu Shenghuo, et al. Regionlets for generic object detection [C]//Computer Vision (ICCV), 2013 IEEE International Conference on IEEE, 2013: 17-24. (in Chinese)
 - [4] Cheng Mingming, Zhang Ziming, Lin Wenyan, et al. BING: binarized normed gradients for objectness estimation at 300fps [C]//Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on IEEE, 2014. (in Chinese)
 - [5] Gall Juergen, Lempitsky Victor. Class-specific Hough Forests for Object Detection [M]// Decision Forests for Computer Vision and Medical Image Analysis. London, Springer: 2013: 143-157.
 - [6] Gall Juergen, Yao Angela, Razavi Nima, et al. Hough forests for object detection, tracking, and action recognition [J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2011, 33(11): 2188-2202.
 - [7] An Meng, Jiang Zhiguo, Zhao Danpei, et al. Space shape object feature training and detection and recognition method based on Hough forest [J]. *Infrared and Laser Engineering*, 2011, 40(8): 1582-1588. (in Chinese)
- 安萌,姜志国,赵丹培,等. 基于Hough树林的空间有形目标特征训练与检测识别方法 [J]. 红外与激光工程, 2011, 40(8): 1582-1588.