# 利用稀疏化生物视觉特征的多类多视角目标检测方法

杨亚威[1],李俊山[1],杨 威[2],赵方舟[1]

(1. 第二炮兵工程学院,陕西 西安 710025;
2. 第二炮兵指挥学院,湖北 武汉 430012)

**摘 要**:鉴于生物视觉特征在目标分类中的良好性能,采用了一种基于稀疏化生物视觉特征的多类多视角目标检测方法。首先利用特征稀疏化方法对生物视觉的标准模型进行改进,有效提高了目标的可分性,然后利用滑动窗口的方法构建基于稀疏化生物视觉特征的目标检测器,采用局部邻域抑制算法完成特定目标的检测任务,最后通过构建场景中待检测目标的词典,对每类目标分别设计滑动检测器以完成多类多视角目标的检测任务,实验结果表明,该方法具有很好的检测性能。

**关键词**:生物视觉; 稀疏特征; 目标检测; 滑动窗口
**中图分类号**:TP391 **文献标志码**:A **文章编号**:1007-2276(2012)01-0267-06

# Multiclass and multiview object detection approach based on sparse biological vision features

Yang Yawei[1], Li Junshan[1], Yang Wei[2], Zhao Fangzhou[1]

(1. The Second Artillery Engineering University, Xi′an 710025, China;
2. The Second Artillery Commanding University, Wuhan 430012, China)

**Abstract:** As the biological vision features shows superior performance on object classification, a multiclass and multiview object detection approach based on sparse biological vision features was adopted. Firstly, the standard model of biological vision was improved with the technique of sparse features, which improved the separability of object effectively. Then, the object detector based on sparse biological vision features was designed with the technique of sliding window, and the detection task was completed via local neighborhood suppression algorithm. At last, the multiclass and multiview object detection task was accomplished through building object dictionary and designing several object detectors in the scene. The experimental results show that the proposed approach exhibits a robust performance.

**Key words:** biological vision; sparse features; object detection; sliding window

**收稿日期**:2011-05-10; **修订日期**:2010-06-13
**基金项目**:国家自然科学基金(60772151)
**作者简介**:杨亚威(1986-),男,硕士生,主要研究方向为基于生物视觉的目标识别与跟踪。Email:13484468703@163.com
**导师简介**:李俊山(1956-),男,教授,博士生导师,博士,主要研究方向为图像处理与目标识别、网络信息安全、电子对抗模拟与仿真。Email:lijunshan403@163.com

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

## 0 引 言

计算机视觉系统对于场景中的某些目标(例如人脸、车辆等)已经具有很好的检测与识别能力。尽管场景中的目标可能出现视角、光照、背景等的变化,以及目标遮挡等其他不确定性因素,一些计算机视觉系统[1-3]仍能很好地检测并识别出目标。但是与生物视觉系统相比,计算机视觉系统的泛化能力仍然存在着很大的不足。生物视觉系统可以很好地检测识别出复杂场景下多类多视角的目标,而计算机视觉系统则相差甚远。因此,从生物视觉系统中寻找启发以改进现有计算机视觉系统是一个很好的途径。

复杂背景下的目标检测问题通常被看作是一个二值分类问题,即区分目标类与背景类。目前主要有以下两种检测方法:一是将分类器转换成一个滑动检测器,通过在图像(或者图像缩放后的金字塔集)上滑动该检测器来对每个局部窗体进行分类[3-5];二是通过兴趣点检测器来提取目标的一些兴趣点,然后对兴趣点位置上具有特定尺寸的局部窗口进行分类[6]。无论采用何种检测方法,一个快速且具有较高精度的分类器是必不可少的,目前常用的分类器有支持向量机[3]、朴素贝叶斯[7]、高斯混合[6]、boosting 决策树[2]等。尽管目前的这些目标检测方法都取得了很好的效果,但它们的工作主要涉及到检测某一类目标,大部分集中在人脸和车辆。为了解决多类多视角目标的检测问题,选取鲁棒性好的目标特征和分类器是必不可少的,参考文献[8-10]已经很好地完成了这方面的工作,但这些文献中的目标都是独立于背景的,未做复杂背景下目标的分割与检测工作。

在目标检测领域,好的目标特征比好的分类器更加重要。与早期的灰度值、小波系数、RGB 直方图等目标特征相比,近年来提出的目标特征,如 Lowe 的尺度不变特征转换(SIFT)[11],Dalal 的方向梯度直方图(HoG)特征[12],Poggio 的标准模型特征(SMFs)[13]等都取得了更好的效果。其中,SMFs 是利用生物视觉的标准模型(ST)提取的,ST 模型以生物神经学原理为基础,模拟灵长类动物视皮层感受野的认知过程,所得 SMFs 具有很强的表征能力,对于标准图像库的识别精度优于 SIFT 和 HoG[14-15]。

文中以生物视觉 ST 模型为基础,利用特征稀疏化方法获取目标的稀疏化 SMFs。对目前常用的目标检测与识别方法进行研究,构建了基于 SMFs 的目标检测器来检测场景中的特定目标。通过构建场景中的多目标词典并以此来建立不同的滑动检测器,进而完成多类多视角的目标检测任务。实验结果表明,文中方法具有很好的检测性能,在对麻省理工学院生物和计算学习中心(MIT CBCL)街景库的目标检测实验中取得了很好的效果。

## 1 特征提取

### 1.1 生物视觉 ST 模型

Hubel 等人在视皮层 17 区第 4 层上发现了对特殊朝向的条形光刺激有强烈反应的包括简单细胞和复杂细胞的感受野构型,并提出了通过汇集操作完成从简单细胞到复杂细胞的过程。研究表明:简单细胞的基本性质是对空间频率和朝向的选择性,并具有特征检测的功能,利用多通道 Gabor 滤波进行特征提取是其神经学方面的基础。复杂细胞拥有更大的感受野,通过汇集操作来完成对简单细胞的响应。在基本的视皮层理论中,简单单元通过寻找感受野中的首选刺激来计算特征,复杂单元通过汇集局部简单单元来构建视觉通道。这种交替的简单-复杂和滤波-汇集过程重复进行,构建了增强的不变性特征,这些特征同时对于增强的复杂刺激具有选择性。在计算机的实现中,这个最终特征被表示成一个监督的学习算法[16]。Poggio 等人在前人神经学研究的基础上构建了基于生物视皮层感受野认知的标准模型,利用该模型提取的 SMFs 在目标分类中具有很好的鲁棒性。

ST 模型主要由 4 层计算单元(S1,C1,S2,C2)组成,其中 S 单元代表简单细胞,利用 Gabor 滤波计算一些高层次的特征;C 单元代表复杂细胞,主要通过一个汇集操作来获取 S 单元的极值,所得特征具有位置和尺度的不变性[14]。ST 模型的层次结构如图1 所示。

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

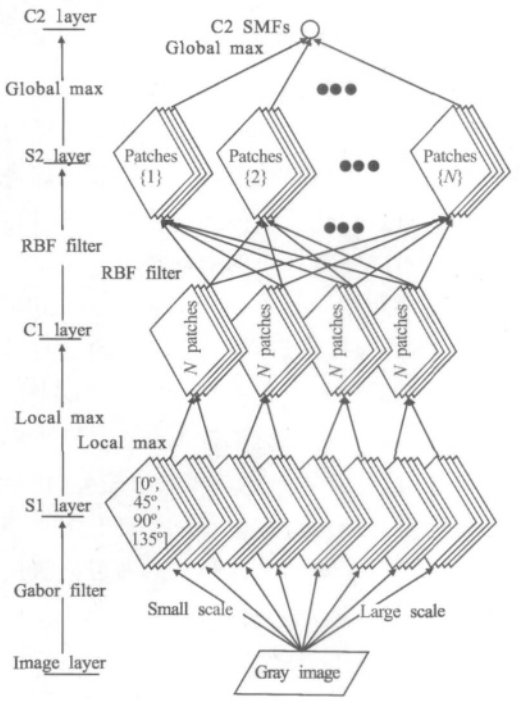


图 1 ST 模型层次结构图

Fig.1 Overview of ST model hierarchy

利用 ST 模型提取 C2 SMFs 特征的具体步骤如下。

(1) 提取 S1 SMFs。利用公式(1)的 Gabor 滤波与输入图像进行卷积,将所得特征在 band 序列上进行排列得到 S1 SMFs。

$$F(x,y)=\exp\left(-\frac{(x_0^2+\gamma y_0^2)}{2\sigma^2}\right)\times\cos\left(\frac{2\pi}{\lambda}x_0\right) \tag{1}$$

式中:$x_0=x\cos\theta+y\sin\theta$;$y_0=-x\sin\theta+y\cos\theta$;$\theta\in[0,\pi]$;$\gamma$ 为 $x_0$ 与 $y_0$ 的比例因子;$\sigma$ 为有效宽度;$\lambda$ 为波长。

(2) 提取 C1 SMFs。首先对 S1 SMFs 从尺度和位置上模拟其在视网膜上的定位情形,通过汇集操作得出局部极大值,然后进行重采样将所得特征在 cPatches 序列上排列得到 C1 SMFs。

(3) 提取 S2 SMFs。利用径向基函数(RBF)滤波求 S2 SMFs,RBF 公式为:

$$r=\exp(-\beta\parallel X-P_i\parallel^2) \tag{2}$$

式中:$\beta$ 为调制强度系数;$X$ 为 C1 单元的 cPatches 特征;$P_i$ 为训练集求得的某一尺度的 cPatches 特征;$r$ 为求得的 S2 SMFs。

(4) 提取 C2 SMFs。对 S2 SMFs 在尺度和位置上通过汇集操作得出全局极大值,得到包含 $N$ 个元素的向量,$N$ 的大小取决于 cPatches 特征的维数。

### 1.2 SMFs 的稀疏化

计算机神经学领域,Olshause 和 Filed 在 1996 年通过研究发现,在对自然图像的统计过程中,增加稀疏性对于生物学不确定性模型的学习起到了非常重要的作用。同样,在机器学习领域,Figueiredo (2003 年)和 Krishnapuram(2005 年)研究发现,增加基函数的稀疏性对于提高分类器的泛化能力起了非常重要的作用[10]。

文中利用特征稀疏化方法,对模型中几个不确定方法进行了重新定义,通过极值定位获取重要方向来减低特征的方向维度,稀疏化了 ST 模型中 S2 单元的输入,提高了目标的可分性。

在原始模型中,S2 单元对 C1 单元中的所有输入都做出响应,即对 C1 cPatches 中每个位置都进行 4 个方向的 Gabor 滤波,然后通过比较和重采样来获取所有位置上 4 个方向的局部极值。但对于实际的神经元,它们很可能更加重视感兴趣的输入。为了稀疏化 S2 单元的输入,把 C1 单元中每个 cPatches 位置上对 S2 响应的数目减少到 1。在特征学习阶段,记下 C1 单元 cPatches 上每个 $n\times n$ 位置上的重要方向(对于 C1 单元的极大响应)的特征和重要性。这样对于 4×4 的 cPatches 样本的结果为 16 个值,而不是原始模型的 64 个值,当计算被稀疏化的 S2 特征时,仍然使用公式(2)的 RBF 滤波,但却只使用了较小的维度:S2 特征仅仅关心每个 cPatches 位置上有重要方向的 C1 输入。这使 S2 单元对于局部混乱特征具有较小的敏感性,提高了泛化能力。由于该方法只关注一个特别重要的方向,而不是原始的 4 个方向,所以能够精确表示方向将是非常重要的。在 C1 和 S1 阶段的处理过程中,把原模型中 Gabor 滤波的 4 方向提高到了 16 个方向,这更好地符合了视觉细胞对于方向的感知精度要远高于 π/4 的目标(Hubell&Wiesel,1959 年)。稀疏化的 C2 SMFs 与原始 C2 SMFs 的比较如图 2、图 3 所示,标注中所选的 6 类目标来自 Caltech 101 库。可以看出,稀疏化 SMFs 比 SMFs 具有更好的可分性。

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

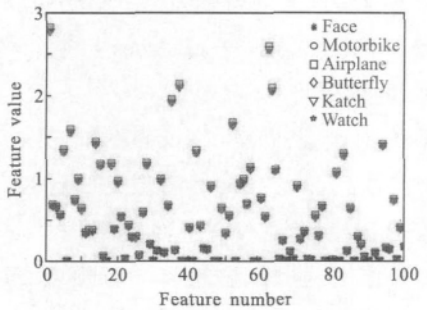


图 2 原始的C2 SMFs 的可分性

Fig.2 Separability of original C2 SMFs

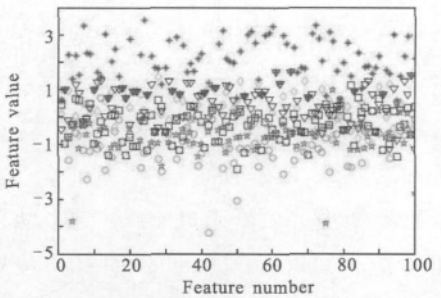


图 3 稀疏化的 SMFs 的可分性

Fig.3 Separability of sparse SMFs

## 2 目标检测

### 2.1 目标检测方法

文中采用经典的滑动窗口的方法[3,5]进行目标检测，其检测流程如图 4 所示。其中训练阶段的任务是训练用于区分目标类与背景类的二值分类器，首先搜集待检测目标的正负样本，其中正样本为场景中包含完整目标的最小图像，负样本为场景中与目标正样本重叠小于重叠阈值的任意背景图像。计算正负样本的稀疏化 SMFs，并以此特征作为输入来训练 SMFs 分类器。测试阶段完成场景中目标的检测任务，采用了经典的滑动窗口方法和标准的局部邻域抑制算法对目标进行检测。由于场景中目标的尺寸不同，仅利用固定大小的滑动窗口去检测目标是不够的，需要用多个尺寸的滑动窗口和 SMFs 分类器在场景图像上进行目标与背景的分类。对于那些被分类器认为是非目标的窗口中心点，映射为 0；那些被分类器认为是目标的窗口中心点，则映射为分类器的信任值。这样将产生一个映射矩阵，该矩阵在分类器认为是目标的窗口中心位置将产生一个最大的映射值。通过局部邻域抑制算法即可检测出目标的具体位置，其具体步骤如下。

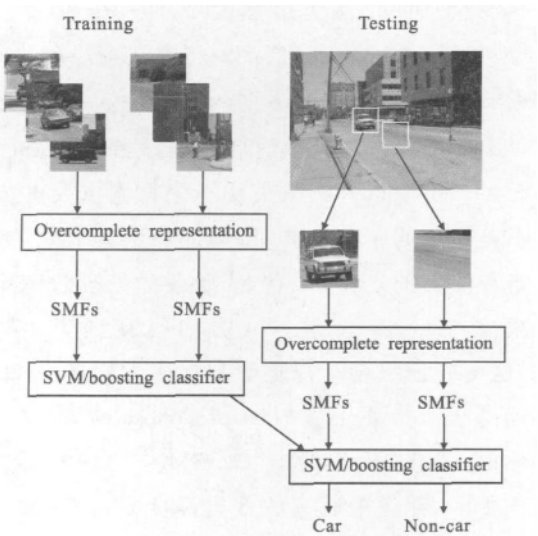


图 4 基于 SMFs 的目标检测流程图

Fig.4 Flow chart of object detection based on SMFs

(1) 把映射矩阵中的所有的非 0 值标记为“非抑制”。

(2) 寻找映射矩阵中最大的非抑制映射值，并把其邻域内的所有映射值设为“抑制”，其中邻域的大小依赖于滑动窗口的尺寸和滑动步长。最大非抑制映射值对应的窗口将被作为检测的目标窗口，邻域内所有映射值均标记为“抑制”则意味着它们对应的目标窗口不再是检测的候选区域。

(3) 重复执行步骤 2 直到映射中的所有映射值为 0 或者处于“抑制”状态。

### 2.2 多类多视角目标检测

检测多类多视角目标主要包括以下 3 个阶段：

(1) 构建目标词典 搜集背景中的待检测目标，构建目标词典。

(2) 训练分类器 对于词典中的每类目标搜集其正负样本，并计算它们的 SMFs，以此作为监督分类算法的输入，训练得出不同目标的分类器。

(3) 检测目标 利用上述训练得到的多个目标分类器和 2.1 节的检测方法，来完成不同目标的检测任务。

## 3 实 验

文中在 Pentium(R) Dual-Core 2.5 GHz CPU，4 G 内存的硬件环境和 Windows XP，Matlab2008 和 OAO SVM 工具(Vojtech & Vaclav, 1999-2003)的软

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

件环境条件下进行了实验。选取 CBCL 的街景库作为测试场景,该库包含 3 457 张图像,图像中包含行人、汽车、自行车 3 类待检测目标,其中每类目标在场景中的位置均已被手动标记并写入脚本,便于研究者对自己的算法进行验证评估。采用正方形的滑动窗口,边长分别为 20,30,40,……,100 像素,相应的滑动步长为 2,3,4,……,10 像素,采用几何中心法对局部邻域抑制算法生成的映射图确定目标位置,对于某些场景的检测结果如图 5 所示。

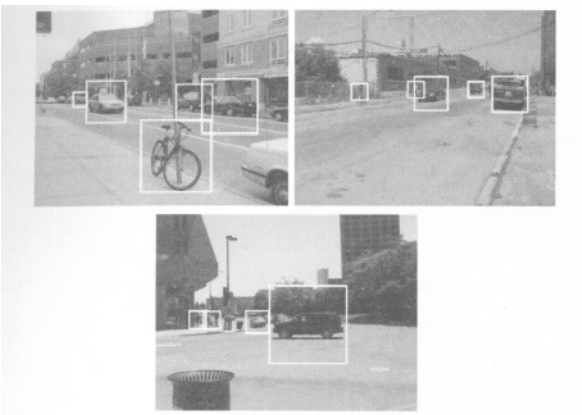

图 5 目标检测实例

Fig.5 Examples of object detection

由图 5 可知,文中的检测方法取得了较好的检测效果,对场景中具有明显形状特征的多视角多形态目标具有很好的检测能力,但对场景中的小目标、部分重叠或者遮掩的目标的检测能力较差。为了定量地评估文中检测方法的性能,采用对预分割图像进行目标分类的方法来近似地评估文中检测方法的性能。通过 LabelMe 工具对街景库中的每类目标进行了标注,用包含目标的最小封闭多边形表示,为了更加近似地评估文中检测方法的性能,首先需要得出封闭多边形的最小外接正方形,即文中方法中的滑动窗口,然后分割出框内目标,对于超出场景边界的目标进行舍弃,在街景库中一共得到 4 055 个汽车、1 141 个行人和 160 个自行车的目标图像,在获取每个目标的同时,利用相同大小的窗口在场景中随机分割出一个背景目标,即负目标,其中正负目标的重叠阈值设为 10%,同样得到 4 055 个非汽车、1 141 个非行人和 160 个非自行车的负目标图像,最后利用分类器对正负目标进行分类,若能正确分类则近似等价于正确检测,这样该分类方法的性能也就近似反映了文中检测方法的性能。

对于每对正负目标的分类实验,采用自抽样法进行验证评估,具体步骤为分别在正负目标中随机抽取 $m$ 个样本用于训练,剩余的样本用于测试验证,这样随机执行 $n$ 次,对分类结果求平均。实验设置 $m \in \{5,10,15,\cdots\cdots,50\}$,$n=10$,每类目标的正确识别率(TPR)和错误识别率(FPR)如图 6 所示。

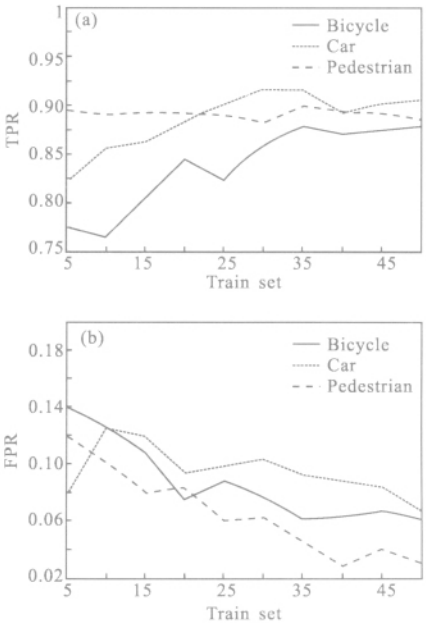


图 6 目标的分类结果

Fig.6 Results of object classification

可以看出,稀疏化生物视觉特征对于多类多视角目标的检测可以达到比较高的精度,在 50 个训练集的条件下,利用 SVM 分类器可以达到低于 8% 的虚警率和近 90% 的正确检测率(针对 SMFs 为高维度向量的特点,SVM 采用线性函数作为核函数)。文中方法和其他一些检测方法的比较如表 1 所示。

**表 1 几种检测方法的比较**

**Tab.1 Comparison of several detection algorithm**

| Detection approach | Car TPR | Pedestrian TPR | Bicycle TPR |
|---|---|---|---|
| Leung(2004)[17] | 75.4% | – | – |
| Fergus et al(2003)[6] | 84.8% | – | – |
| Dalal et al(2005)[12] | – | 86.5% | – |
| Proposed approach | 90.5% | 89.1% | 87.8% |

(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

## 4 结 论

利用稀疏化生物视觉特征作为目标表征，采用滑动窗口和局部邻域抑制算法完成了目标的检测任务，取得了很好的效果，但仍然存在以下问题需要进一步的深化和研究。

(1) 如何进一步提高检测方法的精度和效率。由于滑动窗口方法需要分割出大量的待分类目标，并且计算目标的 SMFs 复杂度比较高，使得文中目标的检测方法效率比较低，下一步拟采用一些兴趣点或兴趣区域的检测方法来进行目标的检测。

(2) 如何更好地检测具有部分重叠或遮掩的目标。由于这些目标丢失了大量的轮廓和形状信息，滑动窗口的方法很难完成该类目标的检测任务，以后的研究中拟采用一些纹理目标的检测方法来完成该类目标的检测任务。

## 参考文献：


[1] Lowe D G. Object recogniti on from local scale-invariant features [C]// International Conference on Computer Vision, 1999: 1150-1157.

[2] Yang Shuishan, He Yonghui, Zhao Wansheng. Strip steel surface defect recognition based on Boosting optimized decision tree[J]. ***Infrared and Laser Engineering***, 2010, 39(5): 954-958. (in Chinese)
杨水山, 何永辉, 赵万生. Boosting 优化决策树的带钢表面缺陷识别技术[J]. 红外与激光工程, 2010, 39(5): 954-958.

[3] Papageorgiou C, Poggio T. A trainable system for object detection[J]. ***Intl J Computer Vision***, 2000, 38(1): 15-33.

[4] Heisele B, Serre T, Mukherjee S, et al. Feature reduction and hierarchy of classifiers for fast object detection in video images [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2001: 18-24.

[5] Wu Yanru, Cheng Yongmei, Zhao Yongqiang, et al. Detection of infrared targets based on Adaboost by feature extraction using KPCA [J]. ***Infrared and Laser Engineering***, 2011, 40 (2): 338-343. (in Chinese)
吴燕茹, 程咏梅, 赵永强, 等. 利用 KPCA 特征提取的 Adaboost 红外目标检测[J]. 红外与激光工程, 2011, 40(2): 338-343.

[6] Fergus R, Perona P, Zisserman A. Object class recognition by unsupervised scale-invariant learning[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2003: 264-271.

[7] Schneiderman H, Kanade T. A statistical model for 3D object detection applied to faces and cars [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2000: 746-751.

[8] Leibe B, Schiele B. Analyzing appearance and contour based methods for object categorization [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2003: 409-415.

[9] Ranzato M A, Huang F J, Boureau Y, et al. Unsupervised learning of invariant feature hierarchies with applications to object recognition [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.

[10] Mutch J, Lowe D G. Multiclass object recognition with sparse, localized features [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2006: 11-18.

[11] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. ***International Journal of Computer Vision***, 2004, 60(2): 91-100.

[12] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2005: 886-893.

[13] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex [J]. ***Nature America***, 1999, 2 (11): 1019-1025.

[14] Serre T, Wolf L, Poggio T. Object recognition with features inspired by visual cortex [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2005: 994-1000.

[15] Serre T, Wolf L. Robust object recognition with cortex-like mechanisms[J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2007, 29(3): 411-426.

[16] Serre T, Kouh M, Cadieu C, et al. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex [R]. US: Massachusetts Institute of Technology, 2005.

[17] Leung B. Component-based car detection in street scene images[D]. US: Massachusetts Institute of Technology, 2004.


(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net