

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2013 2
Vol.18 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

2013 年 2 月 第 18 卷 第 2 期(总第 202 期)

目次

计算机视觉前沿论坛

- 计算机视觉——探索行为理解,认知内心世界 徐光祐,刘允才,章毓晋(131)
- 跨学科前沿与应用的交汇点:日常活动理解 徐光祐,陶霖密,邸慧军(132)
- 时空行为理解 章毓晋(141)
- 视觉机制研究对机器视觉的启发示例 李雄,刘允才(152)
- 视点无关的行为识别综述 冯家更,肖俊(157)

图像处理和编码

- 基于演化算法的卷曲 DCT 图像压缩 李康顺,韦蕴珊,张文生(169)
- 基于欧氏距离图的图像边缘检测 张闯,王婷婷,孙冬娇,葛益娴,常建华(176)
- 基于暗原色先验模型的快速去雾算法 张冰冰,戴声奎,孙万源(184)

图像分析和识别

- Gabor 特征集结合判别式字典学习的稀疏表示图像识别 胡正平,徐波,白洋(189)
- 半监督 k 近邻分类方法 陈日新,朱明早(195)
- 最小距离鉴别投影及其在人脸识别中的应用 黄璞,唐振民(201)
- 图像 LSB 匹配隐藏的预分类检测模型 曹卫权,韩杰思,王宏霞(207)
- 具有旋转鲁棒性的纹理谱描述子 陈刚,陈晓云(214)

计算机图形学

- 利用形体特征的铅笔素描画生成 莫晓斐,丁友东(219)

医学图像处理

- 统计相似度特征的医学图像分割 郭艳蓉,蒋建国,郝世杰,詹曙,李鸿(225)

遥感图像处理

- 结合遗传算法和蚁群算法的高光谱图像波段选择 王立国,魏芳洁(235)
- 相似度自适应样本块的高分辨 SAR 图像阴影修复 赵昊,张弓(243)

-
- 第八届国际数字地球会议征稿通知 封 3

Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 18 No. 2 February 2013

Contents

Forum: Forefront of Computer Vision

Confluence of interdisciplinary frontiers and real-world application: understanding activities of daily living

..... Xu Guangyou, Tao Linmi, Di Huijun(132)

Understanding spatial-temporal behaviors Zhang Yujin(141)

The motivation of visual mechanisms to machine vision: examples Li Xiong, Liu Yuncai(152)

View-invariant action recognition: a survey Feng Jiageng, Xiao jun(157)

Image Processing and Coding

Warped DCT image compression based on an evolutionary algorithm Li Kangshun, Wei Yunshan, Zhang Wensheng(169)

Image edge detection based on the Euclidean distance graph

..... Zhang Chuang, Wang Tingting, Sun Dongjiao, Ge Yixian, Chang Jianhua(176)

Fast image haze-removal algorithm based on the prior dark-channel

..... Zhang Bingbing, Dai Shengkui, Sun Wanyuan(184)

Image Analysis and Recognition

Sparse representation for image recognition based on Gabor feature set and discriminative dictionary learning

..... Hu Zhengping, Xu Bo, Bai Yang(189)

Semi-supervised k -nearest neighbor classification method Chen Rixing, Zhu Minghan(195)

Minimum-distance discriminant projection and its application to face recognition Huang Pu, Tang Zhenmin(201)

Using pre-classification to attack LSB matching Cao Wei-quan, Han Jiesi, Wang Hongxia(207)

Robust rotation-invariant texture spectrum descriptor Chen Gang, Chen Xiaoyun(214)

Computer Graphics

Pencil sketch generation with shape features Mo Xiaofei, Ding Youdong(219)

Medical Image Processing

Medical image segmentation based on statistical similarity feature

..... Guo Yanrong, Jiang Jianguo, Hao Shijie, Zhan Shu, Li Hong(225)

Remote Sensing Image Processing

Band selection for hyperspectral imagery based on combination of genetic algorithm and ant colony algorithm

..... Wang Liguang, Wei Fangjie(235)

Adaptive shadow inpainting from high-resolution SAR image based on exemplar's similarity Zhao Hao, Zhang Gong(243)

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2013)02-152-05

论文引用格式: 李雄, 刘允才. 视觉机制研究对机器视觉的启发示例[J]. 中国图象图形学报, 2013, 18(2): 152-156.

视觉机制研究对机器视觉的启发示例

李雄, 刘允才

上海交通大学自动化系, 上海 200240

摘要: 研究灵长类的视觉系统机制并以此为基础设计机器视觉的算法已成为重要研究方向, 并对机器视觉产生了重要的推动作用。本文从视觉机制和机器视觉方法的角度出发, 分析了两大类视觉机制或模型, 并列列举受其影响和推动的多种重要机器视觉方法: 1) 合作学习和竞争学习机制, 其中合作学习和竞争学习模型相关的机器视觉算法包括立体视觉算法、神经网络、稀疏编码; 2) 简单细胞和复杂细胞模型, 相关的机器视觉算法包括 HMAX 特征、SIFT 描述子和 deep belief network。

关键词: 灵长类动物的视觉机制; 机器视觉方法; 合作学习与竞争学习; 简单细胞与复杂细胞

The motivation of visual mechanisms to machine vision: examples

Li Xiong, Liu Yuncui

Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: It has been a promising methodology that designs machine vision algorithms based on the vision mechanism of primate. In this paper, from the intersection points of vision mechanism and machine vision algorithms, we summarize two categories of important vision mechanisms or models, and present their corresponding machine vision algorithms. 1) Cooperative learning and competitive learning: machine vision algorithms motivated by the models typically include stereo vision, neural networks and sparse coding. 2) Simple cell and complex cell: machine vision algorithms corresponding to the models focus on HMAX feature, SIFT feature and deep belief networks.

Key words: vision mechanism of primate; machine vision algorithm; cooperative learning and competitive learning; simple cell and complex cell

0 引言

目前的机器视觉研究主要包括自底向上和自上而下的两种研究方法。自底向上的方法注重研究视觉的机制, 主要以灵长类动物的视觉机制研究成果为基础设计机器视觉算法。其中最直接的方式是对视觉机制进行建模。这种方法的优点在于, 能直接利用视觉机制的研究成果, 得到简单但非常有效的视觉算法, 并且能直接推广到多种视觉问题。同时,

其主要缺点在于, 许多视觉机制的成果对机制的研究不完善或不全面, 在应用到具体问题上时, 效果很可能不如针对该具体问题设计的算法。自上而下的方法以解决具体问题为目标, 根据具体问题的特点设计相应的算法。其主要优点是, 根据具体问题设计通常较为有效, 目前的大部分机器视觉方法均属于这类。其主要不足在于, 这些方法依赖于要解决的具体问题, 通常难以直接推广到其他问题。

目前国际上有多组研究组从事结合视觉机制与机器视觉算法的研究。普林斯顿大学的 Hopfield 教

收稿日期: 2012-10-08; 修回日期: 2012-12-04

基金项目: 国家重点基础研究发展计划(973) 基金项目(2011CB302203); 国家自然科学基金项目(60833009, 60975012)

第一作者简介: 李雄(1982—), 男, 2008 年上海交通大学模式识别与智能系统专业博士研究生, 主要研究方向为统计学习、计算机视觉和神经计算, 研究成果发表于 CVPR、ACM MM 和 AAAI 等。E-mail: lixiong@sjtu.edu.cn

授、布朗大学的 Mumford 教授、Weizmann 的 Ullman 教授、麻省理工的 Freeman 教授、纽约大学的 Simoncelli 教授、RIKEN 的 Fukushima 教授均在该领域做出重要的贡献。麻省理工的 Poggio 教授研究图像表示和识别机制。其发展的计算模型对机器视觉算法产生了基础性的影响。UBC 的 Lowe 教授致力于图像局部特征的研究,涉及视觉神经的接受域(receptive field)特性、简单细胞(simple cell)和复杂细胞理论。其提出的 SIFT 特征^[1]已成为图像表示、匹配和识别领域的里程碑。UCLA 的 Yuille 和朱松纯教授借鉴视觉系统的计算结构,研究发展了层次构成模型(hierarchical composition model),为复杂图像的表示、解析与理解提供了计算框架。其研究主要涉及简单细胞和复杂细胞的特性以及两者之间的相互关系。加州伯克利分校的 Olshausen 教授发展了学习简单细胞接受域响应特征的方法,即机器视觉中具有重要影响的稀疏编码^[2]。

本文首先介绍视觉机制研究的主要实验手段。然后总结了两大种视觉机制和模型:合作学习和竞争学习模型,以及简单细胞和复杂细胞模型。并分别介绍受这些模型推动发展的机器视觉算法。其中合作学习和竞争学习模型相关的机器视觉算法包括立体视觉算法^[3]、神经网络、稀疏编码^[2]。简单细胞和复杂细胞模型相关的机器视觉算法包括 HMAX 特征^[4]、SIFT 描述子^[1]和 Deep Belief Network^[5-6]。值得注意的是,这两类机制之间是相互关联的。例如,简单细胞的接受域特征可通过竞争学习机制从自然图像中得到^[2]。

1 主要实验手段

视觉机制研究主要借助于电极(electrode 或 multi-electrode array)、脑电图(EEG)和核磁共振图像(fMRI)等。其中电极可以植入清醒的或麻醉的灵长类动物的视觉皮层中,图 1(a)所示为用猕猴进行实验的实验装置;图 1(b)为电极植入大脑的示意图。植入电极时根据所测电信号对视觉刺激的响应特性来调整所测视觉区域,例如通过神经接收域的变化来确定 V1 和 V2 区域。电极能精确地测量 1~100 个以上神经元对特定视觉刺激的响应,来估计单个或一组视觉神经的响应特性。EEG 可以从头皮层测量大量神经元(如 10 万)的混合信号,如图 1(c)所示。但是,由于传感器无法

植入视觉皮层的特定区域,难以研究特定的视觉神经元。核磁共振图像可以测大脑区域的活动情况,如图 1(d)所示。

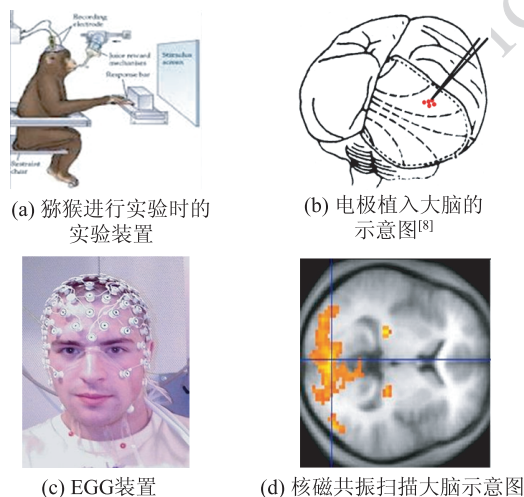


图1 3种与视觉机制研究相关的实验手段

Fig. 1 Three typical experimental methods

2 合作学习和竞争学习

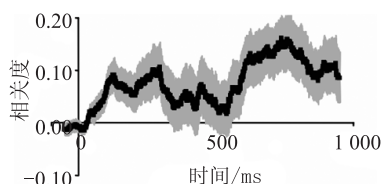
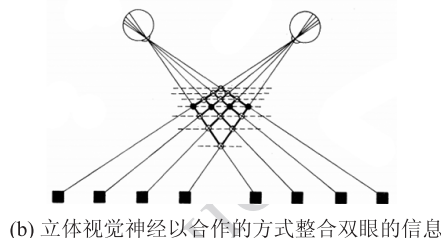
根据神经元之间的作用和连接强度的学习方式,分为合作学习和竞争学习。这两种学习机制为生理学实验所证实。两神经元之间长期的相互激励强化突触的连接强度(合作学习),而长期的压抑降低突触的连接强度(竞争学习)。合作学习通常描述为 Hebbian 学习规则,相关的实例包括立体视觉、神经网络。竞争学习则通常描述为 Anti-Hebbian 学习规则^[7],实例主要有稀疏编码。

2.1 立体视觉

视觉皮层存在主要处理立体信号的立体神经元(disparity tuned neuron)。立体神经元之间主要通过合作式相互作用的方式工作。如图 2(a)所示,在立体神经元对立体信号响应的有效时间内,神经元之间的相关性都几乎是正的^[8]。如图 2(b)所示,立体神经元以合作的方式整合双眼的信息^[3]。这是目前机器视觉中立体视觉算法的基础。

2.2 神经网络

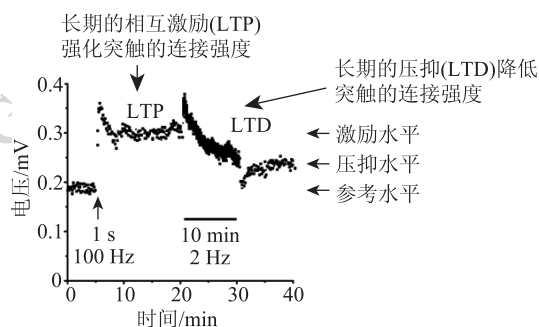
神经网络在机器视觉等问题中有较多的应用。其基础是神经元之间的合作学习机制。图 3(a)是通过电极植入测量得到的神经元连接强度。如图 3(a)所示,长期的相互激励(LTP)强化突触的连

(a) 立体视觉神经在相关性基本为正^[8]

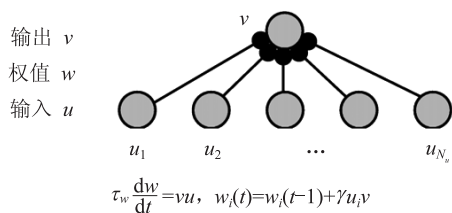
(b) 立体视觉神经以合作的方式整合双眼的信息

图 2 合作式学习与立体视觉

Fig. 2 Cooperative learning and stereo vision



(a) 神经元之间的增强与抑制



(b) 神经网络以合作学习为基础

图 3 神经元作用与神经网络

Fig. 3 Neuronal interaction and neural network

接强度。即两神经元之间的连接强度与他们响应的相关性强度成正比(Hebbian 学习规则)。图 3(b)为神经网络的单元示意图,神经网络的连接权值 w 随着两神经信号 u 和 v 的相关性强度而增强。

2.3 稀疏编码

竞争学习是神经的另一种相互机制,通常描述为 Anti-Hebbian 学习规则^[7]。当一组相互竞争的神经元表达同一视觉信号时,相互抑制对方的响应强度,即激活和使用最少的神经元来表达视觉信号。或者说和以最小的能量来表达视觉信号。如图 4(a)所示,当输出突触 y_1 和 y_2 对应的两神经元同

时响应时,它们之间的连接强度会因竞争产生抑制而降低。

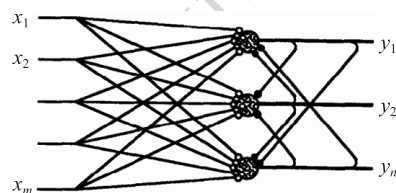
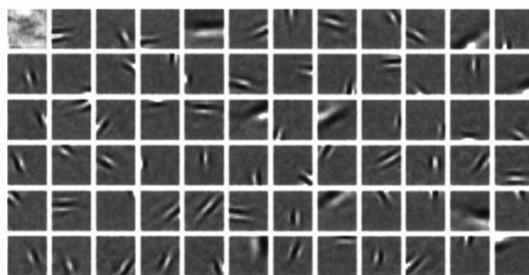
(a) 神经反馈抑制结构^[7](b) 稀疏编码从自然图像中学习到的基^[2]

图 4 竞争学习示意

Fig. 4 Competitive learning

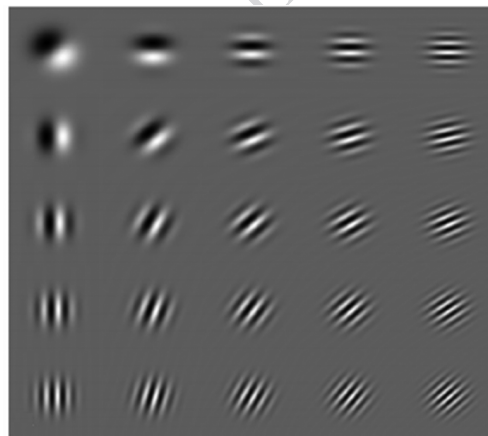
除 Anti-Hebbian 学习规则之外,竞争学习还可以用稀疏编码表示^[2]。即从数据中学习一组过完备基,但仅用很少的基来表达一个视觉信号。稀疏编码通过对组合系数施加 L_1 范数或 Laplace 先验来实现。特别地,稀疏编码从自然图像中学习的过完备基(如图 4(b)所示)与 Gabor 滤波器非常相似,主要编码了方向信息。由于稀疏编码所学习的基通常是少量数据样本信号的线性组合,与图像原型相近,因此对噪声具有很强的鲁棒性,能有效地表达视觉信息。稀疏编码与 SIFT 特征结合是目前图像识别中最有效的方法之一。

3 简单细胞和复杂细胞模型

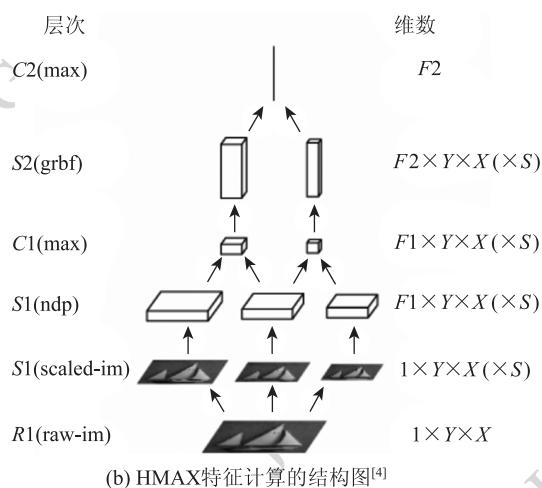
根据神经元处理信号的方式,可以大致分为两类,简单细胞^[1]和复杂细胞。实验发现简单细胞通常只对特定方向和尺度的信号感兴趣。这一性质称为方向选择性(orientation selectivity)^[9]。用一组具有方向选择性的神经元对信号进行处理时,等价于把信号根据尺度和方向进行分解为多个通道,每个通道只编码特定成分的信息。简单细胞的这一特性可以通过 Gabor 滤波器模拟^[10],如图 5(a)所示。

复杂细胞处理信息的方式与简单细胞不同,它将来自不同细胞和通道的信息进行综合和压缩。这

种处理信号方式可能增强输出信号对噪声的鲁棒性。可通过 MAX 运算(如 HMAX^[4])或直方图(如 SIFT^[1])进行模拟。



(a) Gabor滤波器^[10]



(b) HMAX特征计算的结构图^[4]

图5 HMAX示意图

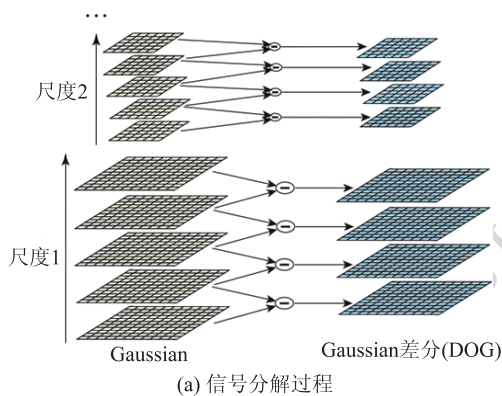
Fig. 5 Illustration of HMAX

3.1 HMAX 特征

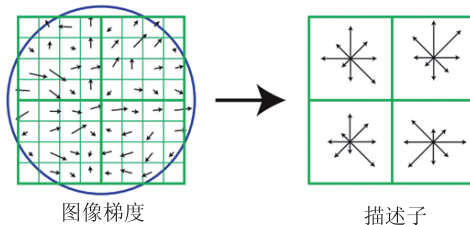
HMAX 是层次式的特征提取框架^[4,11]。以前面所述的简单细胞模型和复杂细胞模型为基础组合而成,结果如图 5(b)所示。包括 S1、C1、S2、C2 4 个层次。其中 S 表示简单细胞, C 表示复杂细胞。首先由 S1 层通过卷积运算将信息分解为多个方向和多个尺度的信号;然后由 C1 层通过 MAX 运算对 S1 的输出信号进行综合和压缩;接着由 S2 层对 C1 的输出进行第二次分解;最后由 C2 层对 S2 的输出进行第 2 次综合和压缩。这种方式较好平衡了简单细胞的选择性和复杂细胞的鲁棒性。文献[12]提出了一种 HMAX 的变体,通过引用神经元间的相互抑制关系对 HMAX 进行了稀疏化。

3.2 SIFT 特征

SIFT 特征^[1]的研究受 Poggio 关于层次化特征计算模型^[11]的启发。首先通过 DoG (difference of Gaussian)将信息分解为多个尺度和梯度方向的信号,然后统计这些编码了梯度方向和尺度信号的分布,分别如图 6(a)(b)所示。这两步分别对应简单细胞和复杂细胞的特性,具体描述可参见文献[1]。SIFT 特征是目前最广泛应用的图像特征之一。与稀疏编码结合,是目前图像识别中效果最好的方法之一。SIFT 特征具有很多种变体^[13]。这些变体主要通过引入新的特性以解决不同的视觉问题,例如 Color SIFT^[13]。但在基本原理上,这些变体仍然与 SIFT 一致,符合简单细胞与复杂细胞的基本特性。



(a) 信号分解过程



(b) 信号综合过程

图6 SIFT 计算示意图^[1]

Fig. 6 Illustration of SIFT^[1]

3.3 卷积深度信念网络(CDBN)

深度信息网络(DBN)由 Hinton 在 2006 年提出^[5],发表于 Science 上。DBN 是种层次式的特征计算模型,每层的性质和结构均类似。CDBN^[6]是种基于 DBN 结构的层次式特征学习模型,结构如图 7(a)所示。与 DBN 的相比,CDBN 引入了汇集层(pooling layer),包括检测层(detection layer)和汇集层两种不同的层。检测层用一组基将输入信号分解为多个通道,汇集层对检测层的输出进行综合和压缩。检测层和汇集层的行为与作用分别对于简单细胞和复杂细胞。这类无监督的特征学习模型,在多

种视觉识别应用中达到了很好的效果。CDBN 在人脸表示和合成中应用的结果如图 7(b)所示。

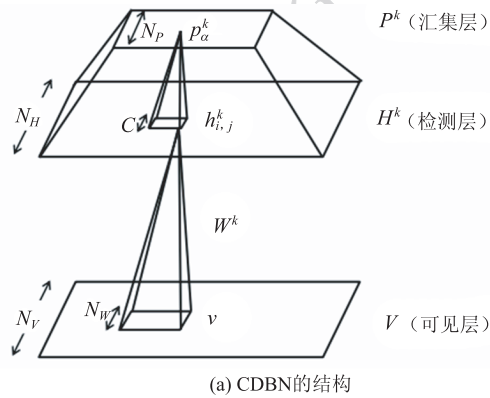


图7 CDBN 结构和结果示意^[6]

Fig. 7 The structure and results of CDBN^[6]

4 结 语

简单介绍了视觉机制中的两类理论和模型:合作学习与竞争学习机制、简单细胞与复杂细胞机制。并分别介绍了受它们启发而产生的机器视觉方法。这些方法广泛应于视觉的各类问题中。与以具体问题为目标设计的方法相比,这些方法具有很好的适应性和推广能力。这些实例说明,从视觉机制出发研究和解决机器视觉中的具体问题这一思路具有良好的前景,并可能对视觉研究产生持久而深远的影响。但由于视觉机制研究中存在大量待探索的问题,并且每一步的发展都非常艰难,因此将视觉机制

的研究成果转化为机器视觉方法需要克服很多困难。对研究人员而言,也存在风险和挑战。

参考文献 (References)

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 2(60): 91-110.
- [2] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. Nature, 1996, 381(13), 607 - 609.
- [3] Marr D, Poggio T. Cooperative computation of stereo disparity [J]. Science, 1976, 194(4262): 283-287.
- [4] Serre T, Wolf L, Poggio T. Object recognition with features inspired by visual cortex [C]// Proceedings of CVPR. New York: IEEE, 2005: 1-7.
- [5] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786): 504-507.
- [6] Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations [C]//Proceedings of ICML. Montreal, Canada; ACM, 2009:1-8.
- [7] Foldiak P, Földiák P. Forming sparse representations by local anti-Hebbian learning [J]. Biol. Cybern., 1990, 64(2): 165-170.
- [8] Samonds J M, Potetz B R, Lee T S. Cooperative and competitive interactions facilitate stereo computations in macaque primary visual cortex [J]. The Journal of Neuroscience, 2009, 29(50): 15780-15795.
- [9] Hubel D H, Wiesel T N. Receptive fields of single neurons in the cat's striate cortex [J]. J. Physiol. 1959, 148(3): 574-591.
- [10] Lee T S. Image representation using 2D Gabor wavelets [J]. IEEE Transactions of Pattern Analysis and Machine Intelligence, 1996, 18(10), 959-971.
- [11] Riesenhuber M, Poggio T. Models of object recognition [J]. Nature Neuroscience, 2000, 3(11): 1199-1204.
- [12] Mutch J, Lowe D. Object class recognition and localization using sparse features with limited receptive fields [J]. International Journal of Computer Vision, 2008, 80 (1): 45-57.
- [13] Sande K E A, Gevers T, Snoek C G M. Evaluating color descriptors for object and scene recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32 (9): 1582-1596.