

基于 GaborSIFT+NNScSPM 图像特征抽取算法研究

江爱文¹ 王春恒² 肖柏华²

摘要 视觉信息的特征表示是计算机视觉场景图像理解研究中的核心内容. 基于 GaborSIFT+NNScSPM 的图像特征抽取算法, 借鉴生物视觉机制中的相关研究成果, 有机结合了 HMAX 层次计算模型的思想和非负稀疏编码的策略, 较为合理地模拟了生物视觉皮层中视觉处理的过程. 在 15 类场景图像和 Caltech101 两个公开数据集上进行了实验验证, 实验结果表明我们所提出的算法较同期算法有着良好的分类性能.

关键词 特征抽取, 生物视觉机制, HMAX, 非负稀疏编码, 语义分类

DOI 10.3724/SP.J.1004.2011.01183

An Image Feature Extraction Method Based on GaborSIFT+NNScSPM

JIANG Ai-Wen¹ WANG Chun-Heng² XIAO Bai-Hua²

Abstract Feature representation of visual information is one of core research topics in computer vision and image understanding. In this paper, we propose a feature extraction method based on GaborSIFT+NNScSPM, trying to combine HMAX model with non-negative sparse coding to mimic the information process in V1 area in visual cortex. We have test our proposed method on two public data sets (15 scenes and Caltech101), and the experiment results show that our method outperforms the existing ones.

Key words Feature extraction, biological inspired feature, HMAX, non-negative sparse coding, semantic classification

场景图像理解是在对场景图像的各种处理和解析的基础上, 解释场景图像的内容及相关高层语义知识, 这个过程是视觉信息和知识信息的处理分析过程, 通常也称为图像解释和场景分析. 近年来, 基于视觉计算的场景信息语义理解已经成为国际上研究和应用的热点之一. 视觉信息的特征表示、图像目标识别和场景语义分类是研究的核心内容.

当前研究者提出了大量的特征表示形式, 用来表示图像描述的内容. 其中主流的特征表示策略可以归纳为以下四大类形式:

1) 基于低层次描述的特征表示. 这类方法直接对场景图像的颜色、纹理等属性进行描述, 然后用这些表示来对图像进行分类. 主要的方式可细分为全局式和子块划分式. a) 全局式: 直接抽取整幅图像的低层特征, 用于图像分类. Vailaya 等^[1] 在颜色信息方面, 利用图像在 LUV、HSV 颜色空间上的空间矩、直方图、一致性向量等; 在边缘信息方面, 利

用图像的边缘方向直方图、边缘方向一致性向量等. Chang 等^[2] 采用颜色和纹理这两种主要特征作为图像的描述. 将像素的 RGB 值转换成颜色波长, 量化形成颜色直方图. 同时采用正交镜像滤波的小波变换抽取图像的纹理分布信息. b) 子块划分式: 首先对图像进行简单的区域块划分, 然后针对每个小区域块进行底层特征抽取、分类, 最后根据每个小区域块的综合情况考虑得到图像的分类结果. 该类方法最早在 1997 年由 Szummer 等提出^[3], 他们对图像子块进行独立分类, 使用多数投票分类器获得最终结果. Serrano 等^[4] 沿用了类似的思想. 不同在于采用支持向量机 (Support vector machine, SVM) 对每个小块进行分类, 经过回归计算可以得到每个小块的分类置信度, 降低分类不准确带来的误差.

2) 基于中间层次语义的特征表示. 所谓中间层语义的概念目前也尚未存在明确的定义. 因此, 大量的研究者提出了不同形式的中间语义表示策略, 主要可以大致分为以下几种: a) 基于语义目标物 (Semantic object) 的中间层表示^[5-7]. 这是最为经典的一类方法. 研究者事先定义一组“语义目标物”集合 (如天空、草地和山脉), 然后场景图像的语义信息定义为由该场景图像中出现的目标物集合构成. 这种方法的瓶颈在于需要有一个初始“语义目标物”检测过程. b) 基于局部语义概念的中间层表示^[8-10]. 试图避免语义目标物的分割和检测过程, 采用更为广义的中间形式的语义表示. 其主体

收稿日期 2010-09-15 录用日期 2011-05-17
Manuscript received September 15, 2010; accepted May 17, 2011

国家自然科学基金 (60835001, 60933010) 资助
Supported by National Natural Science Foundation of China (60835001, 60933010)

1. 江西师范大学计算机与信息工程学院 南昌 330022 2. 中国科学院自动化研究所复杂系统智能控制与管理国家重点实验室 (筹) 北京 100190

1. College of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022 2. State Key Laboratory of Intelligent Control and Management of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

思想是,首先定义一组“视觉词典”,学习某类场景图像中“视觉单词”的分布情况,作为图像的特征表示.这类方法大部分均基于“词袋”(Bag-of-words, BOW)的表示形式.c)基于图像整体语义属性的中间层表示,认为场景图像的语义和场景图像的整体结构布局有关,不需要做图像分割或局部区域处理,突破了传统的观念.这类方法最典型的代表就是 Oliva 和 Torralba 提出的图像空间包络属性(Spatial envelope)^[11].

3) 基于生物视觉机制的特征表示.近年来随着神经生理学、认知科学、计算神经科学等学科的蓬勃发展,大量生理解剖学和心理学的事实为人们对于生物视觉系统的深入理解提供了丰富的实验基础^[12].

在研究人眼视觉系统的图像理解机制时,发现人们能快速理解场景的主旨信息(Gist)^[13-14].这些主旨信息表现为视觉系统对图像空间布局的敏感性.

Poggio 等^[15]结合神经生物学模型和若干假设,提出了一种简单有效的层次化前馈结构模型(Hierarchical Model and X,即 HMAX 模型).

Serre 等^[16]扩展了 HMAX 模型,提出较为明确的四层前馈计算框架.他们的工作进一步细化了 HMAX 模型的各个层次的功能,具体实现了简单细胞层的调节操作和复杂细胞层的最大化(Max)操作,并引入对特征字典学习的思想,使得整个计算模型能够提供各种鲁棒性的特征.

Mutch 等^[17]继而又在 Serre 等的研究工作基础上,采用简单稀疏化策略确定 Serre 模型中 S_2 层各原型区域的主方向,并通过侧抑制(Lateral inhibition)参数来约束 S_1 和 C_1 层非主方向的输出.

生物视觉中的层次计算模型^[15]在一定程度上反映了视觉皮层的结构和功能特点,模拟不同的视觉神经机制和心理学现象,形成有效的视觉场景的表示.目前对于视觉层次计算模型的研究仍处于发展阶段,还存在着许多值得去深入研究和探讨的问题^[18-19],同时可以为解决图像理解的关键问题提供新思路和新方法.

4) 基于稀疏编码的特征表示.稀疏编码理论和模型是神经生物学、计算机科学和心理学的交叉研究领域.随着对生物视觉系统研究的进一步深入,以及计算机技术的飞速发展,近年来稀疏编码理论及应用逐渐成为了人们研究的热点.

Olshausen 等^[20]提出稀疏编码模型,通过定义稀疏性约束来优化学习重构误差最小,得到类似于简单细胞响应特性的基函数.

Lee 等^[21]利用非负矩阵分解(Non-negative matrix factorization, NMF)学习图像的基于部分(Parts-based)的表示.

Hoyer^[22]提出一种非负的稀疏编码(Non-negative sparse coding)模型,根据局部区域重构误差最小,可以有效地模拟视觉皮层神经元感受野的属性.

Yang 等^[23]对密集采样的局部区域 SIFT 特征进行稀疏编码,结合线性空间匹配金字塔策略,实现对场景图像的特征抽取,具备良好的分类性能.

我们在文献[24]的工作基础上,深入研究了生物视觉中的 HMAX 模型以及非负稀疏编码,进行有效结合,提出了一种新的图像特征抽取策略.因此我们的算法总体可以归为基于生物认知机制的特征抽取这一类研究思路.

1 算法描述

本文提出的图像抽取算法既考虑模拟大脑视觉皮层机制中简单细胞响应到复杂细胞响应的视觉处理过程,同时也考虑了场景主旨的空间分布信息.整个特征抽取流程框架图如图 1 所示.

初级视觉皮层中的简单细胞响应有空间位置、方向选择性,以及带通属性,因此多尺度多方向 Gabor 滤波器用来模拟简单细胞的“感受野”响应.

稀疏、过完备的编码表示被用来考虑从简单细胞到复杂细胞处理过程中的空间非线性特性.研究发现^[18],生物视觉系统不是采用严格采样编码方式(Critically sampled code).例如,猫的神经 V1 区,输入神经元与输出神经元数量比大约是 25 : 1,猕猴的比例大约在 50 : 1 的数量级.在某响应表示时,只有大量神经元中的一小部分的神经元被激活,这种方式就是典型的过完备、稀疏表示方式.各输出部分虽然不是保证完全的线性独立,但是它们的过完备性表示策略能够保证其在统计上拥有较高度上的统计独立性,从而降低它们之间的线性依赖性.

1.1 简单细胞响应: 多尺度多方向 Gabor 滤波器

多尺度多方向 Gabor 滤波器的数学表达式为

$$Gabor(x, y | \gamma, \lambda, \sigma) = \exp\left(-\frac{X^2 + \gamma Y^2}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda} X\right)$$

其中, $X = x \cos \theta - y \sin \theta$ 和 $Y = x \sin \theta + y \cos \theta$, x 和 y 介于滤波器窗口大小, θ 表示滤波器方向, $\theta \in [0, \pi]$, γ 表示纵横比 (Aspect ratio), σ 是有效带宽,即尺度信息, λ 是波长.

Gabor 特征具有优良的空间局部性和方向选择性,能够抓住图像局部区域内多个方向的空间频率(尺度)和局部性结构特征,并且对光照等具有一定的鲁棒性.通过调整不同的尺度参数和方向参数,例如采用三种不同尺度和四种不同方向($\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$)的 Gabor 滤波器,可以抽取并

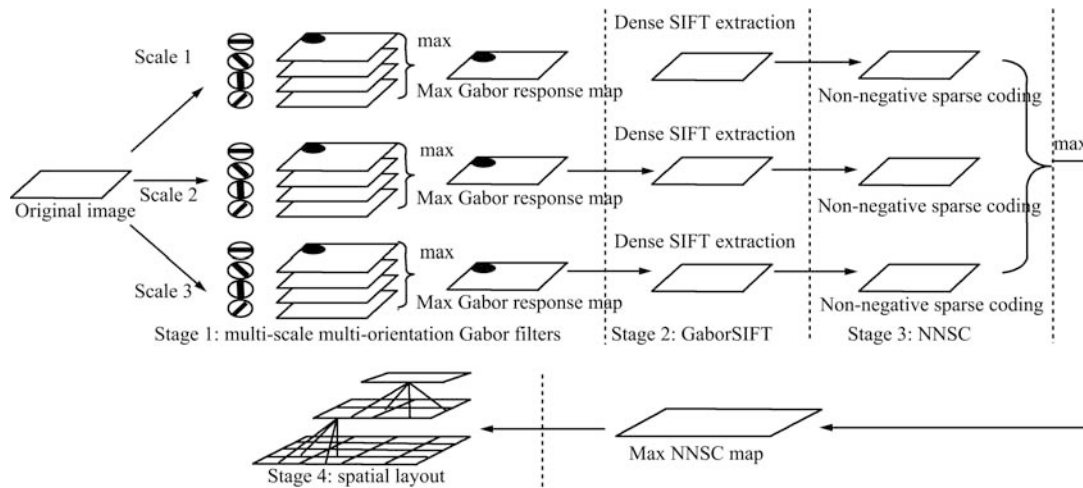


图 1 基于 GaborSIFT+NNScSPM 的特征抽取框架

Fig.1 The framework of the feature extraction algorithm based on GaborSIFT+NNScSPM

用来表示简单细胞感受野的响应特性. 在每一尺度的滤波图像上每一个位置, 记录该位置滤波响应最强的那个方向的响应能量值以及它们相应的方向, 从而得到每个尺度上的最大 Gabor 响应映射图 (Max gabor response map), 如图 2 所示. 这个映射图是稀疏型的, 在每个位置上的响应只对最重要的那个方向敏感, 而不是对所有方向都敏感, 因此更能体现简单细胞的空间位置和方向的选择性. 最大 Gabor 响应映射图的物理意义在于本质上是对滤波结果进行特征选择的过程, 保留图像局部区域最主要的信息, 避免细节干扰信息, 在一定程度上使得特征响应能够克服局部杂乱 (Local clutter) 的干扰, 增强了特征的泛化能力.

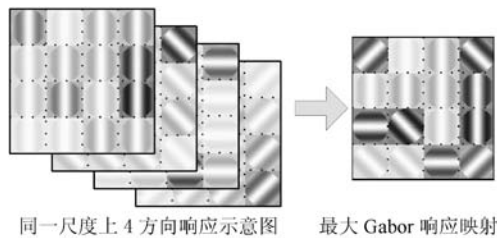


图 2 最大 Gabor 响应映射图

Fig.2 Max Gabor response map

最大 Gabor 响应映射图在每个位置上均具有两个属性 (响应能量值和方向值), 因此我们对该映射图密集抽取大量的局部小块 (Local patch), 计算每个局部小块的 Gabor 能量方向直方图特征, 如图 3 所示. 假设将每个小块分成 4×4 等分, 每一等分内 Gabor 映射存在 4 个方向, 那么该特征的特征维数是 $16 \times 4 = 64$ 维. 因为这里特征的抽取方式 SIFT (Scale-invariant feature transform)^[25] 特征中梯度

方向直方图特征 (Histogram of gradient, HoG) 的抽取方式颇为类似, 并且在不同尺度的最大 Gabor 响应映射图上可以抽取相应局部特征, 因此为了叙述的方便, 我们称该局部特征为 GaborSIFT 特征.

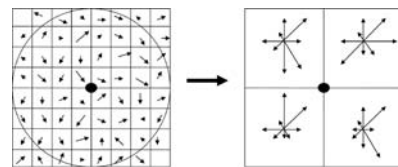


图 3 Gabor 能量方向直方图特征抽取示意图

Fig.3 Histogram of Gabor energy orientation

1.2 复杂细胞响应: GaborSIFT 的非负稀疏编码 (Non-negative sparse coding, NNSC)

Olshausen 等^[20] 阐述了如何在稀疏编码的理论框架下来理解 V1 中简单细胞的感受野的经典特性. 其基本思想是观测数据 x_j 由一组基本模式 (基) a_i 稀疏、线性累加来近似 $x_j = \sum_{i=1}^n a_i s_i$, 其中 s_i 是隐含变量, 具有一定的独立性和稀疏性. 学习得到的基与 Gabor 滤波器的响应有着相似之处. 尽管如此, 他们没有考虑 V1 中的复杂细胞的行为. 复杂细胞响应对应于高一层次的特征抽取. 我们考虑采用稀疏、过完备的编码方式来模拟从简单细胞到复杂细胞处理过程中的空间非线性特性.

另外, V1 中的神经元对背景一般也会有较小刺激响应, 但是刺激不可能为负值. 每个神经元的激励响应均是有正面效用. 在传统的标准稀疏编码中, 隐含变量 (即系数) s_i 可正可负, 当需要进行聚合操作时, 会存在一定的相互抵消的消极作用. 因此在本文所提出的算法中, 我们采用非负的稀疏编码方式来

模拟 V1 中神经元激励行为。

我们注意到简单细胞响应阶段抽取的 Gabor-SIFT 特征向量中元素值均为正值, 因此可以采用非负稀疏编码, 其目标函数为

$$C(A, S) = \frac{1}{2} \|X - AS\|^2 + \lambda \sum_{ij} S_{ij}$$

约束条件: $\forall ij : A_{ij} \geq 0, S_{ij} \geq 0, \forall i : \|\mathbf{a}_i\| = 1, \mathbf{a}_i$ 表示矩阵 A 的第 i 列, $\lambda \geq 0$. 本文的非负稀疏编码算法基于 Hoyer 的工作¹, 具体实现过程如算法 1 所述。

算法 1 非负稀疏编码算法 (NNSC).

Input: Initial A^0 and S^0 by random strictly positive matrices, scale each column of A^0 to unit norm.

Iterate until convergence:

- 1) $A^t = A^t - \mu(A^t S^t - X)(S^t)^T$;
- 2) $A^t = \max\{A^t, 0\}$;
- 3) $A_i^{t+1} = A_i^t / \|A_i^t\|$;
- 4) $S^{t+1} = S^t \cdot ((A^{t+1})^T X) / ((A^{t+1})^T (A^{t+1}) S^t + \lambda)$.

针对每个尺度上最大 Gabor 响应映射图, 我们对抽取的 GaborSIFT 特征进行非负稀疏编码. 从而, 在所有尺度上, 每个 GaborSIFT 对应的位置上均有一个非负稀疏编码表示. 在同一位置上, 对所有尺度上的相应位置上的 GaborSIFT 非负稀疏编码表示进行最大化操作, 参见图 4 所示, $\max NNSC = \max\{nmsc_scale1, nmsc_scale2, nmsc_scale3\}$, 得到一幅最大化稀疏编码映射图 (Max NNSC map). 这种映射蕴含了一定的尺度信息, 同时最大操作模拟实现了简单细胞响应到复杂细胞响应过程中的非线性特性。

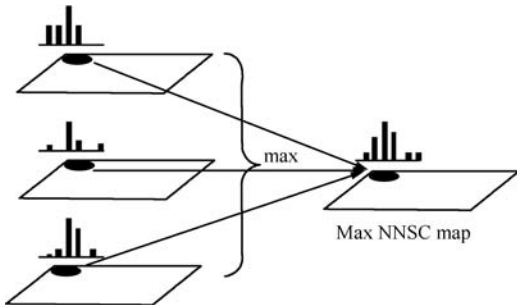


图 4 对稀疏编码表示进行最大化操作示意图

Fig. 4 Maximize operator for sparse representation

采用 GaborSIFT 形式, 而非直接采用 Gabor 滤波后响应作为稀疏编码的输入特征, 是因为考虑到计算的效率. 直接将 Patch 小块作为输入, 将导致特征维数等于 Patch 的大小, 维数较高, 限制了后续稀疏编码算法的效率, 也相应限制了 Patch 大小的

选择自由度. 采用 GaborSIFT 可以对 Patch 信息进行有效压缩, 更为有效灵活, 而且在实验过程中, 实验效果略有提高.

1.3 空间分布信息 (Spatial layout): 空间层次性聚合策略

图像内容的空间分布信息是图像分析过程中所需的一种重要信息, 如图像的全局结构性主旨信息 (Gist). 我们延续了文献 [9] 和文献 [23] 中的策略, 将 maxNNSC 映射图空间粗略划分成 $2^l \times 2^l$ ($l = 0, 1, 2$) 块 (Block). 在每个小块区域中, 对其中的特征点集进行聚合. 本文考虑三种不同的聚合策略。

$$\text{sum : } \mathbf{z}_j = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i$$

$$\text{energy : } \mathbf{z}_j = \sqrt{\frac{1}{N} \sum_{i=1}^N \mathbf{s}_i^2}$$

$$\text{max : } \mathbf{z}_j = \max\{\mathbf{s}_{j1}, \mathbf{s}_{j2}, \dots, \mathbf{s}_{jN}\}$$

其中, N 表示第 j 个 Block 中特征的数目, \mathbf{s}_i 表示每个特征非负稀疏编码后的向量. 每个 Block 中的局部特征经过聚合之后, 形成一个向量, 然后将所有的 Block 的向量归一化, 连接起来形成图像的空间金字塔特征表示形式, 作为图像的最终特征表示形式。

1.4 创新之处: 与 HMAX 和 ScSPM 的区别

本文提出的新的图像特征抽取框架是基于 HMAX 模型和非负稀疏编码的工作基础. 但是又与 HMAX 以及同期工作 ScSPM^[23] 存在着重要的区别。

区别于 HMAX 之处: 1) 在模拟简单细胞响应时, 我们选择性地确定主响应方向, 传统的 HMAX 考虑所有可能的方向, 因此相比之下我们的做法更能体现出方向的选择性; 与 HMAX 的一些后续改进方法 (如 Mutch^[17] 也选择了主方向) 相比, 我们计算主方向的另外一个目的是为了更方便每一位置上的 GaborSIFT 特征的计算, 而 Mutch 是为了简化 S2 特征的输入和计算, 因此目的还是有所区别的. 2) 我们是通过学习的方式, 学习出局部区域的过完备词典, 进而达到局部区域的稀疏编码表示, 而 HMAX (包括后续的改进算法) 的词典获得采用的是随机采样选取方式, 存在着较大的随意性. 3) HMAX 没有利用空间分布信息。

区别于 ScSPM 之处: 1) 我们的算法有着一定生物视觉背景理论支撑, 如 GaborSIFT 的计算过程

¹<http://www.cs.helsinki.fi/u/phoyer/software.html>

以及计算 maxNNScmap; 2) 更重要的不同是, 我们采用的是非负稀疏编码方式, 而 ScSPM 采用的是传统的标准稀疏编码. 一方面, 非负的稀疏编码在模拟神经元激励响应上更为合适^[22]; 另一方面, 在局部区域聚合时, 考虑到每个局部区域的每个响应分量都是有用的假设前提, 非负性能保证做到这一点, 而传统的标准稀疏编码因为系数的可正可负性, 可能会产生相互抵消的消极作用. 这在后面的实验中会继续分析.

2 实验分析

在目前最流行的两个公共数据集 (15 类场景、Caltech101 数据集) 上对我们的算法进行了验证. 为了能够实现比较详细的对比, 我们对采用非负稀疏表示和采用传统稀疏表示这两种不同思路所带来的影响也进行了比较. 1) GaborSIFT+NNScSPM 表示在计算 GaborSIFT 特征基础上, 进行非负稀疏表示 (NNSC), 具体描述如前述; 2) GaborSIFT+ScSPM 表示在计算 GaborSIFT 特征基础上, 进行传统的稀疏表示 (SC). 由于每一个 GaborSIFT 特征在进行传统稀疏表示之后, 系数可正可负, 因此在计算不同尺度间的最大稀疏响应时 (类似 1) 中的 maxNNScmap), 我们采取的策略是最大最小原则 (MaxMin), 即保留系数的绝对值最大项对应的原始值. 采取 MaxMin 策略原则, 主要考虑到稀疏编码表示思想出发点之一是最优线性重构; 保留不同尺度之间对应的正、负绝对响应系数最强部分, 是为了力求不同尺度融合后仍能保留对应部分的主要信息, 不破坏重构的质量; 假使系数简单进行直接最大化, 有可能会使重要的强负响应信息丢失, 造成部分必要信息不完整. 非负稀疏系数 (NNSC) 由于非负性, 不同尺度间可以直接求最大 (效果等同于 MaxMin 原则). 为了使得对比更为全面、客观, 我们对传统稀疏系数, 除了采用 MaxMin 原则, 还对比了直接进行系数最大化处理的结果, 记为 GaborSIFT+ScSPM (max). 这两种思路最后的空间层次聚合策略均同样采用前述的三种策略.

2.1 15 类场景图像数据集

该数据集包含 15 类自然场景图像², 总计 4485 幅图片, 每类图片数目在 200 到 400 之间. 部分样本如图 5 所示.

为了具有可比较性, 我们采用与文献 [9] 和文献 [23] 相同的实验策略, 每类随机选取 100 幅图片, 共计 $15 \times 100 = 1500$ 幅图片作为训练集, 剩下的作为测试集. 我们将图像归一化为 256×256 大小的图

片. Gabor 滤波器尺度大小 $\sigma = \{3, 3.828, 5\}$, 方向 $\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$. 抽取 GaborSIFT 时, 采样的局部区域大小为 16×16 , 稀疏编码的词典大小设为 1024. 分类器同样采用线性分类器. 多类分类器采用“一对多”的方式. 整个过程重复实验 10 次, 计算性能的平均值及方差作为最终的性能结果.

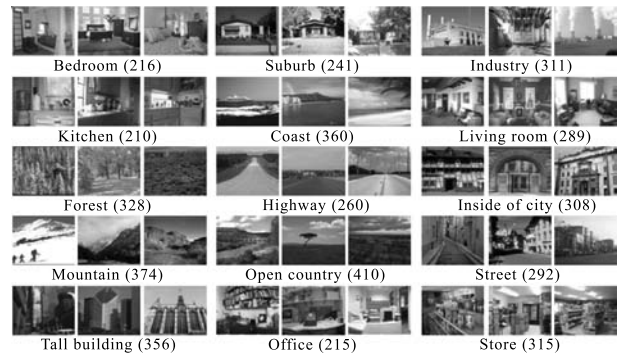


图 5 15 类自然场景图像示意图

Fig. 5 Image samples in 15 scenes

与同期的其他结果进行性能比较, 具体的比较结果如表 1 所示.

表 1 15 类场景数据集上分类性能比较 (%)

Table 1 Classification performances on 15 scenes (%)

数据集	分类性能	
GaborSIFT+NNScSPM	sum	83.37 ± 0.75
	energy	82.70 ± 0.64
	max	80.64 ± 0.91
GaborSIFT+ScSPM	sum	75.40 ± 0.35
	energy	78.26 ± 0.69
	max	79.31 ± 0.54
GaborSIFT+ScSPM(max)	76.50 ± 0.36	
ScSPM ^[23]	80.40 ± 0.45	
KSPM ^[9]	81.40 ± 0.50	

2.2 Caltech101 数据集

Caltech101 数据集³ 总共包含了 101 类不同类型的目标物 (如动物、花、笔记本等). 每类样本个数在 31 到 800 之间. 该数据集的特点是类内差异较大. 部分样本如图 6 所示.

我们随机选择每类的 30 个样本作为训练样本, 余下的样本作为测试样本. 将图像均归一化到 256×256 大小, 其他基本参数设置同 15 类场景数据集实验. 稀疏词典大小设为 1024. 分类器同样采用线性分类器. 多类分类器采用“一对多”的方式.

²http://www-cvr.ai.uiuc.edu/ponce_grp/data/

³http://www.vision.caltech.edu/Image_Datasets/Caltech101/

整个过程重复实验 5 次, 计算性能的平均值及方差作为最终的性能结果. 我们与同期的其他结果进行了性能比较, 具体的比较结果如表 2 所示.



图 6 Caltech101 部分样本

Fig. 6 Image samples in Caltech101

表 2 Caltech101 数据集上分类性能比较 (%)

Table 2 Classification performances on Caltech101 (%)

数据集	聚合策略	分类性能
GaborSIFT+NNScSPM	sum	74.10 ± 0.60
	energy	73.05 ± 0.35
	max	71.20 ± 1.0
GaborSIFT+ScSPM	sum	66.25 ± 0.75
	energy	68.32 ± 0.48
	max	70.60 ± 0.35
GaborSIFT+ScSPM (max)		67.90 ± 0.55
ScSPM ^[23]		73.2 ± 0.54
KSPM ^[9]		64.40 ± 0.80
Jim Mutch ^[17]		56

从在两个不同数据集上的实验结果来看, 我们发现在 GaborSIFT+NNScSPM 的结果中 sum 的聚合策略分类性能是最好的. 与此形成鲜明对比的是, 在 ScSPM 中, 类似于 sum 的 abs 聚合方式结果却不是最优的. 同时我们也发现, 从 GaborSIFT+ScSPM 的结果中可以得出与 ScSPM 类似的结论. 这个现象不难解释, 因为 ScSPM 的初始稀疏编码系数不是非负的, 它们在空间聚合策略阶段的对象是经过人为强制取绝对值转化为非负形式之后的系数. 至于在 GaborSIFT+ScSPM 中, 由于我们 max 的策略采取的是最大最小原则, 效果和绝对值后取最大有类似效果, 只不过聚合之后的特征系数依然有正有负. sum 策略是直接对系数进行相加, 由于系数存在正负值, 在相加的过程会产生一定的抵消. 因此这种情况下, 根据实验结果得出与 ScSPM 基本一致的趋势结论 (即 max 的效果是最好的), 也是合理的. 但是在 ScSPM 中对稀疏表示系数进行人为强制转化显然有不合理之处. 在 GaborSIFT+ScSPM 对可正可负的系数进行直接相加的 sum 聚合策略也不能得到合理的解释. 在 GaborSIFT+NNScSPM 中, 我们直接采用

非负稀疏编码算法得到非负的系数, 这样在此基础上进行后续的一系列聚合操作就显得比较合理而自然, 同时实验的效果也证明了这一点. 相比之下, 基于 GaborSIFT+NNScSPM 的算法思路无论从解释上还是效果上均更具有一定的优势. 但是, 基于 GaborSIFT+NNScSPM 的算法也还存在着一定的不足, 有待改进. 与 ScSPM 算法相比, 在算法效率上, 由于基于 GaborSIFT+NNScSPM 的算法直接采用的是原始非负稀疏编码算法, 没有进行优化, 在效率上制约了算法的速度. 因此, 在将来的工作中, 我们将针对算法效率方面, 对非负稀疏编码算法运算效率做进一步的改进.

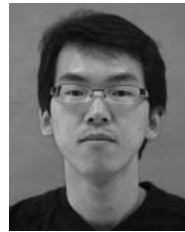
3 结束语

本文提出一种基于 GaborSIFT+NNScSPM 的图像特征抽取算法. 多尺度多方向 Gabor 滤波器模拟简单细胞感受野的响应特性, 抽取 GaborSIFT 特征, 非负稀疏编码的最大化操作 (Max) 考虑简单细胞到复杂细胞之间的非线性. 整个算法系统框架试图有机结合 HMAX 思想和非负稀疏编码思想, 较为合理地模拟生物视觉皮层视觉处理的过程. 我们在 15 类场景及 Caltech101 两个公开数据集上进行了实验验证. 实验结果显示, 我们提出的算法较之同期算法在分类性能上有着良好的优越性. 合理借鉴生物机制进行视觉特征抽取算法研究, 是一个非常有吸引力的研究方向. 我们的工作可以为此提供有意义的参考.

References

- Vailaya A, Figueiredo M A T, Jain A K, Zhang H J. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, 2001, **10**(1): 117–130
- Chang E, Goh K, Sychay G, Wu G. CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, **13**(1): 26–38
- Szummer M, Picard R W. Indoor-outdoor image classification. In: *Proceedings of International Workshop on Content-Based Access of Image and Video Database*. Bombay, India: IEEE, 1998. 42–51
- Serrano N, Savakis A E, Luo J B. Improved scene classification using efficient low-level features and semantic cues. *Pattern Recognition*, 2004, **37**(9): 1773–1784
- Fan J P, Gao Y L, Luo H Z, Xu G Y. Statistical modeling and conceptualization of natural images. *Pattern Recognition*, 2005, **38**(6): 865–885
- Luo J B, Savakis A E, Singhal A. A Bayesian network-based framework for semantic image understanding. *Pattern Recognition*, 2005, **38**(6): 919–934

- 7 Vogel J, Schiele B. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 2007, **72**(2): 133–157
- 8 Li F F, Perona P. A Bayesian hierarchical model for learning natural scene categories. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA: IEEE, 2005. 524–531
- 9 Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA: IEEE, 2006. 2169–2178
- 10 Quelhas P, Monay F, Odobez J M, Gatica-Perez D, Tuytelaars T. A thousand words in a scene. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(9): 1575–1589
- 11 Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 2001, **42**(3): 145–175
- 12 Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 1959, **148**(3): 574–591
- 13 Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, 2006, **155**(1): 23–36
- 14 Li F F, Van Rullen R, Koch C, Perona P. Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 2002, **99**(14): 9596–9601
- 15 Poggio T, Riesenhuber M. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 1999, **2**(11): 1019–1025
- 16 Serre T, Wolf T, Bileschi S, Riesenhuber X, Poggio T. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(3): 411–426
- 17 Mutch J, Lowe D G. Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision*, 2008, **80**(1): 45–57
- 18 Olshausen B A, Field D J. How close are we to understanding V1? *Neural Computation*, 2005, **17**(8): 1665–1699
- 19 Carandini M, Demb J B, Mante V, Toulhurst D J, Dan Y, Olshausen B A, Gallant J L, Rust N C. Do we know what the early visual system does? *Journal of Neuroscience*, 2005, **25**(46): 10577–10597
- 20 Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, **381**(6583): 607–609
- 21 Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, **401**(6755): 788–791
- 22 Hoyer P O. Modeling receptive fields with non-negative sparse coding. *Neurocomputing*, 2003, **52-54**(1): 547–552
- 23 Yang J C, Yu K, Gong Y H, Huang T. Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, USA: IEEE, 2009. 1794–1801
- 24 Jiang A W, Wang C H, Xiao B H. A new biologically inspired feature for scene image classification. In: Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey: IEEE, 2010. 758–761
- 25 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110



江爱文 江西师范大学计算机与信息工程学院讲师。2010 年获中国科学院自动化研究所博士学位。主要研究方向为图像处理与模式识别。本文通信作者。

E-mail: aiwen.jiang@ia.ac.cn

(JIANG Ai-Wen Lecturer at Jiangxi Normal University. He received his Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences in 2010. His research interest covers image processing and pattern recognition. Corresponding author of this paper.)



王春恒 中国科学院自动化研究所研究员。主要研究方向为图像处理与模式识别。E-mail: chunheng.wang@ia.ac.cn

(WANG Chun-Heng Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers image processing and pattern recognition.)



肖柏华 中国科学院自动化研究所研究员。主要研究方向为图像处理与模式识别。E-mail: baihua.xiao@ia.ac.cn

(XIAO Bai-Hua Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers image processing and pattern recognition.)