整合局部特征和滤波器特征的空间金字塔匹配模型

高常鑫,桑 农
(华中科技大学图像识别与人工智能研究所多谱信息处理技术国家级实验室,湖北武汉 430074)

摘 要: 本文提出一种场景分类方法,通过整合局部特征和滤波器特征获得丰富的表征信息,并利用空间金字塔匹配模型提取空间上下文信息.该方法有如下四个特点:(1)通过转换将滤波器很好地嵌入空间金字塔匹配模型中;(2)在滤波器特征转换的过程中,采用降采样和平均操作,在空间密度和空间范围两者之间取得了很好的折衷;(3)将滤波器特征和局部特征组合起来,获得了更强的描述能力;(4)捕获了像素域和调制域的互补信息.同时,在三个数据库上的实验证明了该方法的有效性.

关键词: 基于上下文的表征; 空间金字塔匹配; 像素域; 调制域; 场景分类
中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2011) 09-2034-05

Unifying Local Features and Filterbank Features in the Spatial Pyramid Matching Model

GAO Chang-xin, SANG Nong
(*Institute for Pattern Recognition and Artificial Intelligence, National Key Laboratory of Science & Technology on Multi-spectral Information Processing, Huazhong University of Science and Technology, Wuhan, Hubei* 430074, *China*)

Abstract: This paper presents an approach to scene classification, which unifies local features and filterbank features to capture rich representation information, and extracts spatial context information using the spatial pyramid matching (SPM) model. The proposed method has four characteristics. First, filterbank features are successfully embedded into the SPM model by a transformation method. Second, in the transform process, downsampling and average pooling are used to achieve good balance between spatial density and spatial extent. Third, filterbank features and local features are combined to represent images for more discriminative power. Fourth, the complementary information is extracted in pixel and modulation domains. Promising experimental results on three datasets demonstrate the effectiveness of the proposed method.

Key words: context-based representation; spatial pyramid matching; pixel domain; modulation domain; scence classification

1 引言

场景分类是计算机视觉中一个极具挑战性的课题.场景分类在计算机视觉中的应用非常广泛,例如可以用于移动式遥控装置导航,也可以作为特定目标识别的预处理过程[1~3].在计算机视觉的研究中,最常用的场景分类方法是基于目标的方法.通常场景中有一些固定的明显标志,基于目标的场景分类方法正是通过识别这些标志来识别这个场景的[4~6].但是这些方法还包含一些中间步骤,比如分割、特征组织和目标识别,而这些也是计算机视觉中目前尚未完全解决的关键问题.另一方面,人类视觉系统在场景分类方面性能却非常好,远远超过了计算机视觉系统.近年来的一些心理学实验证明人类视觉对于现实世界场景的识别是从场景的整体结构开始的,甚至不需要知道其中具体包含的目标[7~9].人类在200ms内就可以捕获一幅场景中一定的感知和语义信息(图像和其中部分目标的语义类别以及它们的一些属性),称为场景的"gist"[10,11].考虑到人类视觉的优越性,很多研究者试图从精神物理学和神经生理学中寻找一些启发来填补低层视觉特征和高层语义概念之间的鸿沟.最近一些计算机视觉研究者提出,将一幅图像当作一个整体,然后利用场景空间分布的低维统计编码来获取场景的语义[1,3,12,13],这种编码特征称为基于场景上下文的表征.基于场景上下文的表征方法比基于目标的方法更加鲁棒,因为它不需要基于目标方法中的那些中间步骤.

收稿日期:2010-11-15;修回日期:2011-03-16
基金项目:国家自然科学基金重点资助项目(No.60736010);中国博士后科学基金资助项目(No.20100480902);中央高校基本科研业务资助(No.HUST:2010ZD034)

本文的目的是研究一种更加有效的场景分类方法.到目前为止,场景分类任务中有两类基于场景上下文的方法取得了很好的效果,分别是 gist 模型[1,2,7,14]和空间金字塔匹配(Spatial Pyramid Matching,SPM)模型[12].为了获得更好的判别能力,两种模型都考虑了粗糙的空间信息.Gist 模型和 SPM 模型中使用的特征分别为滤波器特征和局部特征.很多研究都证明了 SPM 模型在场景分类中的性能更加突出[12].本文提出了一种新的场景分类方法,整合了像素域和调制域信息来描述场景图像.该方法有四个特点:

(1)将 AM-FM 特征(滤波器特征)和 SIFT 特征(局部特征)组合,增强了描述子的判别能力;

(2)特征描述是双通道的场景描述方法,AM-FM 特征[15]和 SIFT 特征[16]分别属于调制域和像素域;

(3)将滤波器特征转换为局部特征,并将其非常好地嵌入 SPM 模型中;

(4)在滤波器转换为局部特征的过程中,使用了两个处理,分别为降采样和平均,这样就能很好地获得空间密度和空间范围的平衡.

在三个数据库上测试了本文的方法,分别针对三个任务设计,在 15 类场景数据库上测试针对场景分类任务;在 Caltech-101 数据库[17]上测试针对多类目标识别任务;在 UIUCTex 数据库[18]上测试针对纹理分类任务.这些数据库上的实验结果证明了本文方法的有效性.

2　将滤波器特征嵌入 SPM 模型

2.1　SPM 模型

特征袋(Bag-of-Features,BoF)模型是一种有效的图像表征模型,该模型将图像描述为局部特征的分布或者直方图统计,很多研究[12,17,19]都证明了该模型的突出性能.但是 BoF 模型对于目标形状的描述性不够,因为它丢掉了局部特征之间的空间关系信息.Lazebnik 等[12]提出了一种新的框架将局部特征之间的空间关系加入 BoF 模型中,称为 SPM 模型.该模型描述空间布局的方法如下:首先将图像按照金字塔结构分为一些子区域,然后在每个子区域上利用 BoF 模型计算特征,最后将所有区域的特征组成一个向量.实验证明该模型对于场景分类任务非常有效.

2.2　将滤波器特征嵌入 SPM 模型

滤波器特征描述一幅图像时,保留了所有滤波器的响应,这就包含了足够多的信息.但是如果把滤波器特征直接嵌入到 SPM 模型中,其性能并不好[12].Lazebnik 等人[12]还指出了其原因:相对于局部特征,滤波器特征的密度太高,而且空间范围太小.因此在将滤波器特征转换为局部特征描述的过程中就需要实现以下两点:减小特征的空间密度和增加特征的空间范围.本文提出的框架通过将滤波器特征转换为局部特征的方法,很好地将全局的滤波器特征嵌入到了 SPM 模型中.为了获得采样密度和空间范围之间的折衷,在转换的过程中使用了降采样和平均的处理.降采样的目的是将滤波器的密集描述转换为稀疏描述;每个降采样位置的特征值是周围一定范围内值的平均响应值,这样就利用了邻域的信息,使得描述更加鲁棒.最后,把各个通道上采样位置的特征值组成一个向量就是该采样特征最终的局部特征向量.图 1 给出了将滤波器特征 AM-FM 特征嵌入 SPM 模型的示意图.转换过程可以总结为三个主要步骤:

(1)对于每一个滤波通道,将滤波结果降采样;

(2)每一个采样位置的特征值通过周围响应值的平均值来表示;

(3)对于每一个采样位置,局部描述子就是这个位置在所有滤波通道上的特征值.

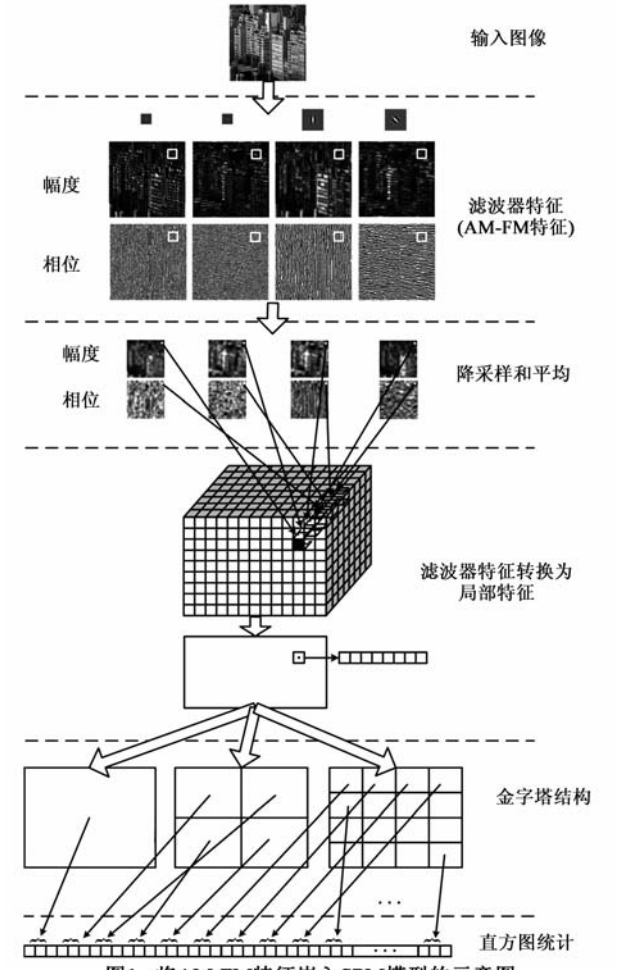

图1　将AM-FM特征嵌入SPM模型的示意图

2.3　整合 SIFT 描述子和 AM-FM 特征的 SPM 模型

SPM 模型中,图像被分成很多局部区域,将这些区域用最匹配的局部特征代替,得到一个局部特征表征的图像;而滤波器特征会保留所有的响应.局部特征是

稀疏的,使得特征获得了很好的不变性.另一方面,滤波器特征则包含了大量的信息.考虑到 SIFT 描述子和 AM-FM 特征的杰出性能,本文提出了一种新的方法,将这两个特征都用于 SPM 模型中.SIFT 描述子[16]是一个非常有效的描述子,它是局部特征,属于像素域;AM-FM 特征[15]则能够很好地描述纹理信息,它是滤波器特征,属于调制域.两种组合的混合特征可以同时捕获像素域和调制域的信息,也可以同时捕获局部特征和纹理特征,因此这个混合特征包含的信息非常丰富.

3 实验结果与分析

本节测试本文提出的基于 SIFT 描述子和 AM-FM 特征的 SPM 模型,并与当前性能最好的方法进行对比.在三个数据库上针对三种任务做测试:在 15 类场景数据库[12]上测试场景分类任务;在 Caltech-101 数据库[17]上测试多类目标识别任务;在 UIUCTex 数据库[18]上测试纹理分类任务.

3.1 实验设置

(1)SPM 模型

在 SPM 模型中,利用 3 层金字塔结果来提取空间信息.这和 Lazebnik 等方法[12]的设置相同,该方法证明了 3 层金字塔结构可以很好地平衡判别能力和泛化能力.降采样和平均是将滤波器特征转换为局部特征过程中的两个步骤,采样间隔设为 8 个像素,平均处理中邻域的大小为 16 × 16 像素.因此,对于一幅 256 × 256 像素的图像,采样后大小为 31 × 31 像素.与 Lazebnik 等方法[12]相同,设字典长度为 200,最后得到的特征向量为 $(4^3-1)\times 200/3 = 4200$ 维.

(2)特征

AM-FM 特征:多成分的 AM-FM 图像模型通过幅度调制来描述局部的对比度特性,通过瞬时频率来描述局部的纹理结构[15].首先用一组 Gabor 滤波器(4 个尺度 8 个方向,共 $4\times 8 = 32$ 个滤波器)分离和解调到各个成分上,一个成分对应一个滤波器通道,即最后将一幅图像分离和解调到 32 个成分上.每个通道都要计算幅度特征和相位特征,也就是说利用 32 个滤波器得到 64 个结果图.

SIFT 描述子:尺度不变特征变换(SIFT)是一种基于直方图的描述子,已经广泛地应用于计算机视觉的多个方面.最初的 SIFT 描述子[16]将图像分为几个不重叠的子区域(网格结构),在每个子区域中统计梯度方向直方图.利用 4×4 的网格结构、8 个方向的 SIFT 描述子来描述该区域,即 SIFT 描述子的维数为 $4\times 4\times 8 = 128$.SIFT 描述子的尺度规定为 16 × 16 的图像区域,这和 Lazebnik 等人[12]的设置相同.

(3)分类器

除此之外,实验中利用支持向量机(SVM)进行训练与测试,针对 SPM 模型中使用了金字塔匹配核(pyramid match kernel)[12].本文的实验中多类分类任务采用了一对多的策略.

3.2 15 类场景数据库

第一组实验是在 15 类场景数据库[12]上测试的.15 类场景数据库包含的图像有:卧室(216 幅)、郊区(241 幅)、工业(311 幅)、厨房(210 幅)、客厅(289 幅)、海岸(360 幅)、森林(328 幅)、公路(260 幅)、市内(308 幅)、山脉(374 幅)、空旷地区(410 幅)、街道(292 幅)、高楼(356 幅)、办公室(215 幅)和商店(315 幅),图像大小约为 300 × 250 像素.图 2 给出了该数据库上一些例子.实验中,每类 100 个样本用于训练,其它的用于测试,表 1 给出了本文的方法与其它一些方法识别率的比较,最好的结果用粗体标出.可以看出,本文的方法性能最好,这就说明将 SIFT 描述子和 AM-FM 特征组合嵌入 SPM 模型中更有利于场景分类任务.图 2 给出了分类结果最好的三类和最差的三类场景,并在相应类别图像和类别名称的下方给出了相应的分类正确率,其中左边是正确率最高的三类场景(郊区、森林和街道),右边是正确率最低的三类场景(卧室、客厅和工业).另外还给出了 15 类场景之间混淆矩阵,如图 3 所示,其中灰度值表示识别率,灰度为 255 表示识别率为 1,灰度为 0 表示识别率为 0,其中第 i 行、第 j 列的格子表示属于类别 i,识别为类 j 的概率.可以看出识别结果和人类的直观感觉很相似,室内场景之间更容易混淆(卧室、客厅和厨房),一些室外场景之间也容易混淆.本节还给出了混淆最严重的 5 组情况,分别为海岸混淆为空旷地区、客厅混淆为卧室、卧室混淆为厨房、工业混淆为高楼、卧室混淆为客厅,观察了数据库中的图像发现,人类同样觉得这些场景类别之间容易混淆.

图2 15类场景数据库中部分样本

表 1 15 类场景数据库上识别率结果比较

方法	本文方法	文献[12]方法	文献[14]方法	文献[17]方法	文献[20]方法
识别率	82.94%	80.55%	75.42%	64.1%	76.7%

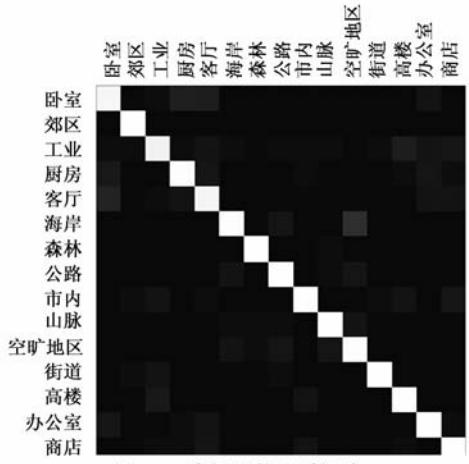

图3　15类场景的混淆矩阵

3.3　Caltech-101 数据库

第二组实验是在 Caltech-101 数据库上测试本文的方法,对应的任务为多类目标识别. Caltech-101 数据库包括 101 类目标,每类包含图像数从 31 到 800 不等.图 4 给出了部分图像.绝大多图像非常清楚,目标的干扰很少甚至没有,而且目标都位于图像中间,甚至目标的姿态都是非常接近的.本文的实验设置和文献[19]相同,即从每类随机选择 30 个作为训练样本,其余的作为测试样本.表 2 给出了多类目标识别的平均识别率,几类方法中最好的结果用粗体标出.为了分析本文提出的方法中两种特征在 SPM 模型中的性能,还给出了在 SPM 模型中只使用 AM-FM 特征的结果.需要说明的是,文献[12]方法是在 SPM 模型中只使用 SIFT 描述子的方法.可以看出本文的方法比其它方法效果都要好.这些实验结果也证实了 Lazebnik 等[12]的观点,即 SPM 模型在目标干扰和姿态变化都很少的情况下性能非常好.而且该情况下 AM-FM 特征在 SPM 框架中比 SIFT 描述子性能好一些,这就说明 AM-FM 特征比 SIFT 描述子更适合这种干扰和姿态变化很小的情况.图 4 中,在相应类别图像和类别名称的下方给出了识别率最好和最差的三类目标,其中左边是最好的三类,右边是最差的三类.可以看出本文的方法同样对于形状信息为主的目标识别率较差,这和 Zhang 等[19]的结论相同.

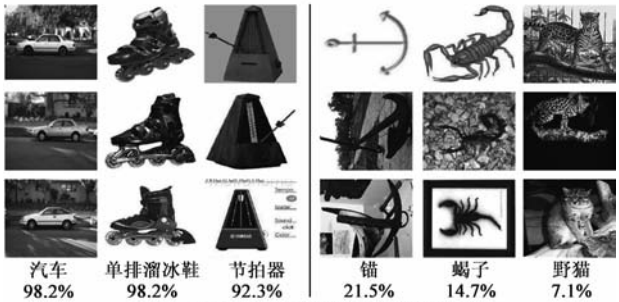

图4　Caltech-101数据库上部分样本

表 2　Caltech-101 数据库上识别率结果比较

方法	本文方法	AM-FM + SPM	文献[12]方法	文献[21]方法	文献[22]方法
识别率	67.10%	66.62%	64.68%	66.43%	56%

3.4　UIUCTex 数据库

最后在 UIUCTex 数据库[18]上测试本文的方法对于纹理分类任务的性能. UIUCTex 数据库包含 25 类纹理类,每类 40 幅图像,图 5 给出了部分图像.该数据库上纹理分类的难点主要体现在以下几个方面,存在非刚体变换、亮度变化和视点变化.实验中每类选择 20 个作为训练样本,另外 20 个作为测试样本,分类正确率结果见表 3.可以看出,对于纹理分类任务,AM-FM 特征比 SIFT 更有效,这是因为 AM-FM 特征能够更有效地描述纹理信息.本文的方法比文献[12]的方法性能好,但是比文献[19]的方法差,其原因可能是本文的方法对于非刚体变换和视点变化太敏感,但是实验结果说明本文的方法在纹理分类方面还是有一定潜力的.图 5 给出了三对最容易混淆的类别,相应类别的正确率在相应类别图像和类别名称的下方给出.可以看出局部结构和纹理频率相似性是导致混淆的主要原因.基于局部特征的方法对于局部特征比较相似而纹理频率不同的情况分类性能比较差,比如 T09&T10,和 T21&T24.本文的方法证明了加入纹理特征对解决该问题有所帮助.

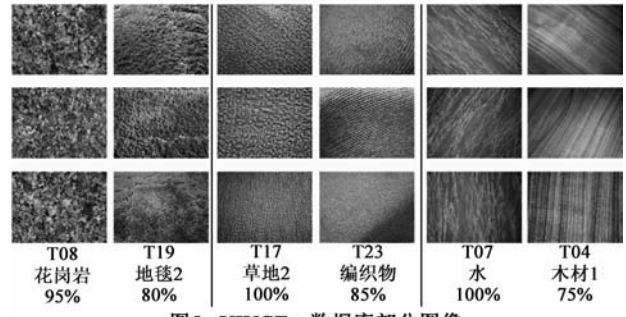

图5　UIUCTex数据库部分图像

表 3　UIUCTex 数据库上识别率结果比较

方法	本文方法	AM-FM + SPM	文献[12]方法	文献[19]方法
识别率	93.72%	90.40%	85.46%	98.5%

4　结论

本文提出了一种新的基于 SPM 框架场景的分类方法,整合了 SIFT 描述子和 AM-FM 特征,充分利用了像素域和调制域的信息.实验证明该方法对于场景分类和目标识别任务非常有效,同时对于纹理分类任务该模型也表现出一定的潜力.但是,从实验结果也可以,看出本文的方法对于目标的几何变化是比较敏感的,为了获得判别能力和泛化能力之间更好的折衷,今后还需要加强特征的几何不变性.采样密度过于密集会影响表征的性能,可以考虑将稀疏表征和密集表征的方法相结合,从而可以提高算法的性能,这也是今后的一个研究方向.

参考文献

[1] Torralba A, Murphy K P, et al. Context-based vision system for place and object recognition[A]. Proceedings of the IEEE International Conference on Computer Vision[C]. Nice, France: IEEE Press, 2003. 273 – 280.

[2] Siagian C, Itti L. Rapid biologically-inspired scene classification using features shared with visual attention[J]. International Journal of Computer Vision, 2007, 29(2): 300 – 312.

[3] 谢昭,高隽. 基于高斯统计模型的场景分类及约束机制新方法[J]. 电子学报, 2009, 37(4): 733 – 738.
XIE Zhao, GAO Jun. A novel method for scene categorization with constraint mechanism based on Gaussian statistical model[J]. Acta Electronica Sinica, 2009, 37(4): 733 – 738. (in Chinese)

[4] Abe Y, Shikano M, et al. Vision based navigation system for autonomous mobile robot with global matching[A]. Proceedings of IEEE International Conference on Robotics and Automation[C]. Detroit, Michigan: IEEE Press, 1999. 1299 – 1304.

[5] Thrun S. Finding landmarks for mobile robot navigation[A]. Proceedings of IEEE International Conference on Robotics and Automation[C]. Leuven, Belgium: IEEE Press, 1998. 958 – 963.

[6] 于林森,张田文. 基于视觉与标注相关信息的图像聚类算法[J]. 电子学报, 2006, 34(7): 1265 – 1269.
YU Lin-sen, ZHANG Tian-wen. Image Clustering based on correlation between visual features and annotations[J]. Acta Electronica Sinica, 2006, 34(7): 1265 – 1269. (in Chinese)

[7] Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42: 145 – 175.

[8] Biederman I, Mezzanotte R J, et al. Scene perception: detecting and judging objects undergoing relational violations[J]. Cognitive Psychology, 1982, 14(2): 143 – 177.

[9] Oliva A, Schyns P G. Diagnostic colors mediate scene recognition[J]. Cognition Psychology, 2000, 41(2): 176 – 210.

[10] Potter M C. Meaning in visual search[J]. Science, 1975, 187(4180): 965 – 966.

[11] Oliva A. Gist of the Scene[A]. Neurobiology of Attention[C]. San Diego, CA: Elsevier Press, 2005. 251 – 256.

[12] Lazebnik S, Schmid C, et al. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE Press, 2006. 2169 – 2178.

[13] 刘硕研,须德,冯松鹤,等. 一种基于上下文语义信息的图像块视觉单词生成算法[J]. 电子学报, 2010, 38(5): 1156 – 1161.
LIU Shuo-yan, XU De, FEN Song-he, et al. A novel visual words definition algorithm of image patch based on contextual semantic information[J]. Acta Electronica Sinica, 2010, 38(5): 1156 – 1161. (in Chinese)

[14] Torralba A. Contextual priming for object detection[J]. International Journal of Computer Vision, 2003, 53(2): 169 – 191.

[15] Kokkinos I, Evangelopoulos G, et al. Texture analysis and segmentation using modulation features, generative models, and weighted curve evolution[J]. IEEE Trans on PAMI, 2009, 31(1): 142 – 157.

[16] Lowe D. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.

[17] Fei-Fei L, Fergus R, et al. Learning generative visual models form few training examples: an incremental Bayesian approach tested on 101 object categories[A]. IEEE CVPR Workshop on Generative-Model Based Vision[C]. Washington: IEEE Press, 2004. 178 – 185.

[18] Lazebnik S, Schmid C, et al. A sparse texture representation using local affine regions[J]. IEEE Trans on PAMI, 2005, 27(8): 1265 – 1278.

[19] Zhang J, Marszalek M, et al. Local features and kernels for classification of texture and object categories: A comprehensive study[J]. International Journal of Computer Vision, 2007, 73(2): 213 – 238.

[20] van Gemert Jan C, Veenman Cor J, et al. Visual word ambiguity[J]. IEEE Trans on PAMI, 2010, 32(7): 1271 – 1283.

[21] Zhang H, Berg A C, et al. SVM-KNN: discriminative nearest neighbor classification for visual category recognition[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE Press, 2006. 2126 – 2136.

[22] Mutch J, Lowe D G. Multiclass object recognition with sparse, localized features[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE Press, 2006. 11 – 18.

作者简介

高常鑫 男,1982 年 7 月出生于山西省平遥县. 现为华中科技大学图像识别与人工智能研究所博士后. 主要研究方向为计算机视觉、模式识别. E-mail: changxin.gao@gmail.com

桑 农 男,1968 年 8 月出生于重庆市. 现为华中科技大学图像识别与人工智能研究所教授、博士生导师. 主要研究方向为图像处理、模式识别、计算机视觉.
E-mail: nsang@hust.edu.cn