

基于 Gist 和 PHOG 特征的场景分类

刘 静, 郭 建, 贺遵亮

(湘潭大学材料与光电物理学院, 湖南 湘潭 411105)

摘 要: 局部 Gist 方法提取的特征维数过高、计算复杂, 单一的 Gist 特征不能很好地描述全局场景。为此, 提出一种将改进的局部 Gist 特征与梯度方向直方图特征进行组合的场景描述方法。采用支持向量机作为分类器, 在 WS 场景库中考察单一特征和组合特征的分类精度, 在 OT 场景库下研究不同数量训练样本对于分类精度的影响。实验结果表明, 与全局 Gist、局部 Gist 等方法相比, 该方法能降低计算的复杂度, 且提高分类正确率。

关键词: 局部 Gist 特征; 梯度方向直方图; 特征组合; 场景描述; 支持向量机; 场景分类

中文引用格式: 刘 静, 郭 建, 贺遵亮. 基于 Gist 和 PHOG 特征的场景分类[J]. 计算机工程, 2015, 41(4): 232-235, 240.

英文引用格式: Liu Jing, Guo Jian, He Zunliang. Scene Classification Based on Gist and PHOG Feature[J]. Computer Engineering, 2015, 41(4): 232-235, 240.

Scene Classification Based on Gist and PHOG Feature

LIU Jing, GUO Jian, HE Zunliang

(Faculty of Materials, Optoelectronics and Physics, Xiangtan University, Xiangtan 411105, China)

[Abstract] In view of complex computation caused by extracting high dimension characteristics with local Gist method, as well as the problem that the sole Gist characteristic can not describe global scenes well, a kind of improved method to describe the scenes is proposed, which combines local Gist characteristics with Histograms of Oriented Gradient (HOG) characteristics. Classification accuracy of the sole characteristics and the combination of characteristics are inspected in the WS scene database using Support Vector Machine (SVM) as the classifier. On this basis, classified precision influenced by different quantity training samples is also studied in the OT scenes database. Experimental results show that this method reduces the computational complexity, and improves the classified accuracy compared with the global Gist, local Gist methods, etc.

[Key words] local Gist feature; Histograms of Oriented Gradient (HOG); feature combination; scene description; Support Vector Machine (SVM); scene classification

DOI: 10.3969/j.issn.1000-3428.2015.04.044

1 概述

场景分类是计算机图片分类的重要分支。将场景图片区分为自然场景与人造场景可以为后续不同类别场景分类提供前期分类保证^[1], 划分成自然场景和人造场景后可以针对这 2 类场景再次提取不同的场景描述特征, 进行第 2 次场景精细划分, 因此, 首先进行自然场景与人造场景的区分有着很重要的实际意义。

全局语义建模方法将场景视为一个整体进行描述, 用全局特征来进行场景描述。当前使用的最好的全局语义建模方法是文献[2-3]提出的采用空间

包络模型作为场景全局语义特征的描述算子 Gist, 将场景图像通过不同方向和不同尺度的 Gabor 滤波器组后划分成 4×4 的网格, 取平均值级联形成图像的全局 Gist 特征。由于全局 Gist 特征是稀疏网格划分的, 因此同一网格中可能包含不同的场景部分, 为此, 文献[4]提出采用局部 Gist 特征作为场景描述算子, 但局部 Gist 特征在进行特征加权的时候只采用了一次加权, 造成了特征维数过高。

梯度方向直方图 (Histograms of Oriented Gradient, HOG)^[5]不仅可以描述局部特征, 同样可以用来描述整幅图像的特征。HOG 特征首先统计细胞单元内的局部梯度方向信息, 然后在若干个细胞单元

作者简介: 刘 静 (1987 -), 男, 硕士研究生, 主研方向: 图像处理, 模式识别; 郭 建, 教授; 贺遵亮, 硕士研究生。

收稿日期: 2014-05-14 **修回日期:** 2014-06-30 **E-mail:** liujing@xtu.edu.cn

组成的区域内进行归一化,而 Gist 特征是对整个图像滤波后在局部区域内取各方向、各尺度的均值。两者都不对图像进行预分割,利用的都是图像在空间域里像素值不连续形成的方向统计信息,将两者组合起来描述场景图像能够提供更丰富的信息,文献[6]将 Gist 特征与 HOG 特征进行组合实现图像的场景分类。

局部 Gist 方法提取的特征维数庞大,HOG 特征没有考虑到图像不同空间尺度划分对分类性能产生的影响。为此,本文提出一种基于 Gist 特征和分层梯度方向直方图(Pyramid Histogram of Oriented Gradient, PHOG)^[7]特征组合的场景分类方法。采用改进局部 Gist 方法替代全局 Gist 方法,用 PHOG 特征替代 HOG 特征,然后进行特征组合实现场景分类。

2 特征提取

2.1 局部 Gist 特征

图像的 Gist 特征提取,在实际的实现中采用的是将图像与不同方向和不同尺度的 Gabor 滤波器组进行滤波,将滤波后得到的图像划分为网格,在每个网格内部取平均值,最后将滤波后得到的所有图像在每个网格均值级联起来,得到图像的 Gist 特征。

一幅大小为 $h \times w$ 的灰度图像 $f(x, y)$ 用 n_c 个通道的滤波器对图像进行卷积滤波,其中, n_c 等于滤波器尺度和方向数的乘积,这样得到 n_c 个滤波后的图像。将每幅滤波后的图像划分为 4×4 的网格,每个网格内取平均值得到一个特征。将每个网格块的特征级联起来,得到图像的 Gist 特征,即:

$$G_i(x, y) = \underset{n_c}{cat}(f(x, y) * g_{mn}(x, y)) \quad (1)$$

其中, $\underset{n_c}{cat}$ 为级联符号; $g_{mn}(x, y)$ 为滤波器组; $*$ 为卷积运算符。

局部 Gist 特征是将上述特征提取过程中的网格划分改为由 16×16 的网格进行划分,对于每一个网格再次采用大小为 4×4 的网格进行划分,将每一个二级网格采取加权求和,权重矩阵 W 构建方法为对应像素到各子区域中心像素归一化距离的倒数,再将所有的二级子网格特征进行级联,得到局部 Gist 特征。该方法细化了场景特征的提取,有效地降低了同一网格中包含不同场景的概率,从而提高了场景分类的正确率,但其还存在一个问题,就是提取的特征维数过高。如一幅大小为 256×256 像素的图像,其提取后的特征维数为: $16 \times 16 \times 4 \times 4 \times 32 = 131\,072$ 维。如果不采取降维处理等手段,特征维数过于庞大,包含很多冗余信息,同时,给数据处理带来很大的困难。

2.2 HOG 特征提取

HOG 特征是一种在计算机视觉中用于目标检测的常用特征描述符,最早是由法国研究机构在 2005 年提出来的,是对图像的局部区域内梯度方向信息的统计。HOG 特征是在一个大小统一的网格密集的细胞单元上计算,且采用重叠部分对比度归一化的方法来提高特征描述的性能,而光照条件变化和局部几何变换这 2 种形变只会出现在更大的空间领域上,所以, HOG 方法对它们可以保持一定程度的不变性。

HOG 特征的提取过程如下:

(1) 对图像进行伽马颜色校正,然后用 2 个一维的模板:

$$\begin{aligned} h_x &= [-1 \quad 0 \quad 1] \\ h_y &= [1 \quad 0 \quad -1]^T \end{aligned}$$

分别对图像的垂直方向和水平方向做卷积运算。在称为细胞的单元内,根据每一个像素的梯度方向,在事先定义的方向区间里,以梯度幅值为权重,统计梯度方向直方图。像素点的梯度方向定义为:

$$G_0(m, n) = \arctan(G_y(m, n)/G_x(m, n)) \quad (2)$$

其中, m, n 分别为像素的行号和列号; G_x, G_y 分别为图像在 x, y 方向的梯度; G_0 为得到的梯度方向。

(2) 在被称为块的区域内对上一步得到的方向直方图进行正规化。用式(3)对块内的直方图进行正规化:

$$H(i) = \frac{H(i)}{\sqrt{\sum_{n=1}^n H(i)^2 + \varepsilon}} \quad (3)$$

其中, $i = 1, 2, \dots, n$, n 为直方图的区间数; ε 为小的正常数。

(3) 使用截断阈值 th 对得到的直方图进行截断:

$$H(i) = \begin{cases} H(i) & \text{if } H(i) \leq th \\ th & \text{if } H(i) > th \end{cases} \quad (4)$$

(4) 用式(3)对截断后的直方图进行正规化,得到最终的梯度方向直方图。正规化后的各细胞直方图按一般的图像扫描方式,即从上到下、从左到右的方式连接起来,形成特征向量。

HOG 特征考虑了图像空间位置的分布,是描述图像形状信息的一种有效方法,但是没有考虑到图像的不同空间尺度划分对分类性能可能产生的影响。

3 特征组合

针对全局 Gist 特征与局部 Gist 的缺点,本文提出改进的局部 Gist 特征。在局部 Gist 特征提取过程中,进行加权以后,每一个二级子块的特征为 16 维,

显然特征维数过大。同时,为了突出中心区域在整个区域所在的比重,将上述 16 维的向量从左到右、从上到下重新排列为 4×4 的矩阵,并再次进行加权,其权重与局部 Gist 特征中所采用的权重相同。这样每一个二级子块就降为一个特征向量,较局部 Gist 特征的维数有了较大幅度的降低,同时,也突出了每一个二级子块的中心区域场景的特征值。具体提取过程如图 1 所示。

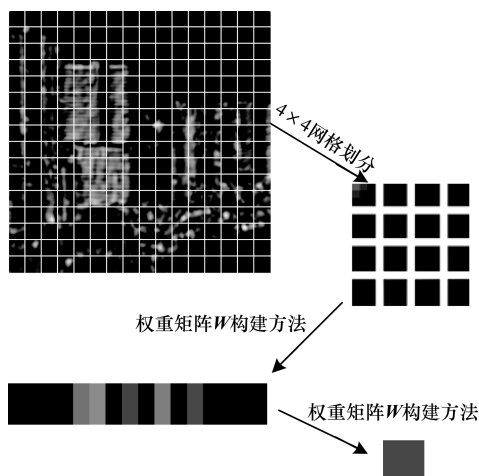


图 1 改进的局部 Gist 方法特征提取过程

图 1 左上角为经过滤波后的图片,将图片采用 16×16 的网格进行划分,经过稠密网格划分后可以发现每一个网格中基本上只包含了一类场景;右上角为每一个一级子网格经过 4×4 的子网格划分后的情况;左下角为第一次采用权重处理后,每一个一级子网格计算出来的特征值;右下角为将一级子网格特征值进行重新排列后再次进行权重处理后得到的特征值。经过这样处理,每一个一级网格提取出一个特征值,大大降低了特征维数,同时,经过权重处理以后,每一子网格的中心部分得以突出。

针对 HOG 特征没有考虑到不同空间尺度划分对分类性能的影响,且其不能很好地描述形状的空间分布,采用 PHOG 特征代替 HOG 特征。PHOG 特征是在 HOG 特征基础上改进得到的。PHOG 提取算法使用空间四叉树分解形成图像的多分辨率表示,将从低分辨率到高分辨率的多级方向梯度直方图级联形成 PHOG 特征。PHOG 表示出图像中物体的局部形状以及形状在空间中的分布,通过统计一个区域内图像边界方向分布情况来表达局部形状,将图像以不同的尺度分割成网格来表达空间布局^[8]。

本文采用将 2 种特征进行首尾相连的方式来进行特征组合^[9],形成一维特征向量,以此来描述场景图片。

4 实验结果及分析

根据上述方法提取图像改进的局部 Gist 特征和 PHOG 特征,并将两者进行组合,得到图像的描述特征。采用支持向量机 (Support Vector Machine, SVM)^[10-11] 对特征向量进行学习,使用 LibSVM^[12] 工具箱,核函数为径向基函数,采用交叉验证得到最优参数,对场景进行自然场景与人造场景进行 2 类分类。每组实验进行 10 次训练与测试,取 10 次的平均值作为测试结果。

本文测试实验的硬件及软件环境:计算机运行内存 4 GB,64 Byte Windows7 操作系统,Matlab2013b。

考虑到 WS 图像库样本数量有限,本文测试只对 OT 库进行训练样本数量与分类正确率的统计,以及特征提取、分类耗时和内存占用率的统计。

4.1 WS 图像库分类测试

WS 图像库是华盛顿大学的场景图像库,本文进行实验测试时选取该库中的 12 个类别,从其中选取合适的图片组合成自然场景与人造场景图像库,其中,自然场景 193 幅,人造场景 153 幅。由于场景数据库中图片数量不多,为充分进行训练,从每类中随机选取一半数量图片进行训练,余下部分作为测试样本。

将本文方法与文献[2]的全局 Gist 方法、文献[4]的局部 Gist 方法、文献[6]的 Gist + HOG 方法进行对比。在实际测试中,文献[6]采用全局 Gist 特征。WS 场景库测试比较结果如表 1 所示。

表 1 WS 场景库下不同方法的分类正确率对比 %

方法	自然场景分类 正确率	人造场景分类 正确率	总体正确率
全局 Gist 方法	88.17	69.23	79.53
局部 Gist 方法	84.95	75.64	80.70
Gist + HOG 方法	87.10	75.64	81.87
本文方法	86.02	79.49	83.04

从表 1 可以看出,在 WS 场景库中,虽然全局 Gist 方法,和 Gist + HOG 方法的自然场景区分正确率比本文方法高,但是总体正确率均低于本文方法,其中,全局 Gist 方法的人造场景正确率不足 70%。

4.2 OT 图像库分类测试

MIT 的 Oliva&Torralba (OT) 场景库共包含了如图 2 所示的 8 类场景,分别为海滩 360 幅、森林 328 幅、山脉 374 幅、乡村 410 幅、高速公路 260 幅、城区 308 幅、街道 292 幅和高楼 356 幅,将其中前 4 类划归为自然场景,后 4 类划归为人造场景,每幅图像大小均为 256×256 像素,共计 2 688 幅。

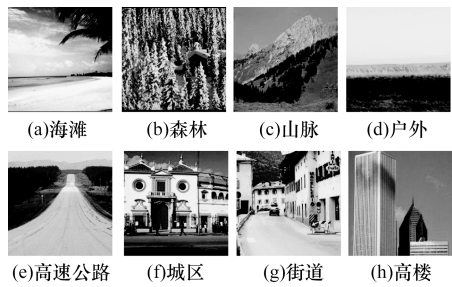


图 2 OT 场景图像数据库

在 OT 场景库下,本文首先比较了不同训练样本数量下 4 种方法分类正确率的变化情况,如图 3 所示。从图中可以看出,随着训练样本数的增加,分类正确率逐步提升,在训练样本数达到 150 左右时,分类正确率基本上趋于稳定,可以看出,本文方法在不同训练样本数量时分类正确率均高于其他 3 种方法。

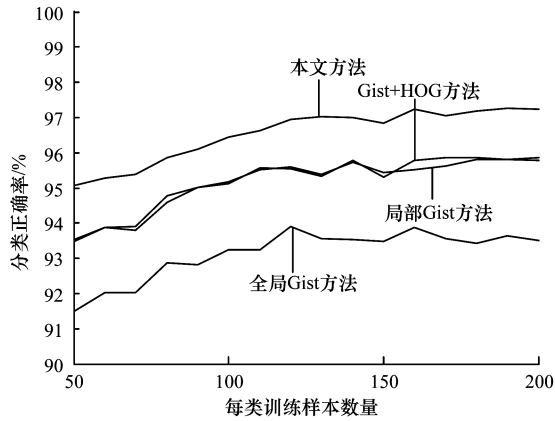


图 3 训练样本数量与分类正确率的关系

综合训练样本数量和测试样本数量考虑,采用训练样本数为 150 时来进行 4 种方法分类正确率的比较。具体比较结果如表 2 所示。

表 2 OT 场景库下不同方法的分类正确率对比 %

方法	自然场景 正确率	人造场景 正确率	总体正确率
全局 Gist 方法	95.76	90.10	93.48
局部 Gist 方法	96.79	93.51	95.43
Gist + HOG 方法	97.13	92.57	95.30
本文方法	97.82	95.45	96.84

从表 2 可以看出,在 OT 场景库中,本文方法能有效提高自然场景与人造场景的区分正确率,正确率达到 96.84%,对于自然场景和人造场景每一类的分类正确率相较于其他方法均有提升。

针对本文实验数据库,还统计了特征提取及分类所占用的时间,以及内存占用情况,具体为:特征提取是将整个库中的图片提取特征后统计每一幅图像特征提取所占用的平均时间;分类时间为统计一

定数量图片分类耗时取平均值得到单幅图像分类耗时;内存占用率是计算机在只运行该程序时统计的物理内存占用率。具体比较结果如表 3 所示。

表 3 OT 场景库下不同方法的性能指标

方法	特征提取 耗时/s	分类 耗时/s	特征提取时 内存 占用率/%	分类时 内存 占用率/%
全局 Gist 方法	0.90	0.000 59	44	44
局部 Gist 方法	2.10	0.060 30	47	60
Gist + HOG 方法	0.93	0.000 62	44	44
本文方法	1.58	0.020 80	45	58

从表 3 可以看出,对于每一张图片从特征提取到分类结束本文方法用时为 1.6 s,虽然较全局 Gist 及 Gist + HOG 2 种方法用时稍长,较速度最快的全局 Gist 方法延时了 0.6 s 左右,考虑到场景分类的应用情况,该时间消耗在可接受范围内。还可以看出,虽然内存占用率在分类时达到了 58%,但在计算机快速发展的今天,相比于本文方法所带来的分类正确率的提升,该限制因素可以忽略不计。因此,从上述分类测试及耗时和内存占用率情况的比较可以看出,在忽略其他因素影响下,本文方法可以达到提升场景分类正确率的目的。

5 结束语

场景分类中局部 Gist 特征维数过高,单一特征不能很好地描述场景。为解决该问题,提出一种基于改进的局部 Gist 特征和 PHOG 特征进行组合的特征向量,以描述图像场景,结合支持向量机进行自然场景与人造场景分类。实验结果表明,该方法的分类正确率高于全局 Gist 方法、局部 Gist 方法和 Gist + HOG 方法。然而,本文只讨论了自然场景与人造场景分类,如何实现 2 类场景中每一类场景的精确区分是今后的研究重点。

参考文献

[1] 刘 静. 基于场景 Gist 的快速场景分类研究[D]. 吉林: 吉林大学, 2013.

[2] 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(3): 145-175.

[3] Oliva A, Torralba A. Building the Gist of a Scene: The Role of Global Image Features in Recognition[J]. Progress in Brain Research: Visual Perception, 2006, 155: 23-36.

[4] 杨 昭, 高 隽, 谢 昭, 等. 局部 Gist 特征匹配核的场景分类[J]. 中国图象图形学报, 2013, 18(3): 264-270.

(下转第 240 页)

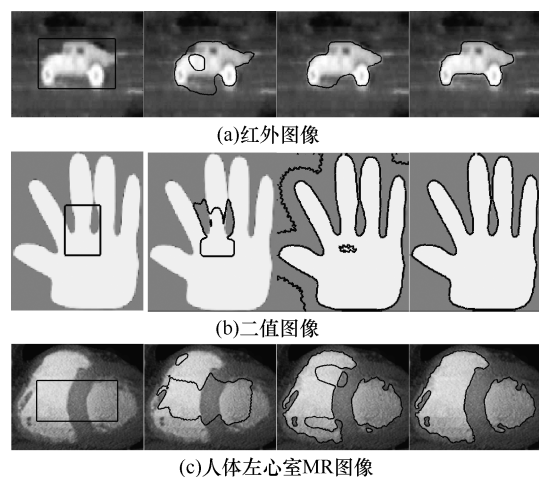


图9 本文模型对其他类型图像的分割效果

5 结束语

针对图像分割中的灰度不均匀和轮廓初始化问题,本文提出一种基于区域的活动轮廓模型。通过融合全局信息和局部信息,既能有效地处理灰度不均匀图像,又能进行灵活的轮廓初始化。实验结果表明,该模型对噪声和弱边缘图像能取得较好的分割效果。今后将考虑将本文模型应用于各种医学图像的分割,并进一步提高模型的性能。

参考文献

- [1] Osher S, Sethian J A. Fronts Propagating with Curvature Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations [J]. Journal of Computational Physics, 1988, 79(1): 12-49.
- [2] Li Chunming, Xu Chenyang, Gui Changfeng, et al. Level Set Evolution Without Re-initialization: A New Variational Formulation [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition.

- San Diego, USA: IEEE Computer Society Press, 2005: 430-436.
- [3] Lie J, Lysaker M, Tai Xuecheng. A Binary Level Set Model and Some Applications to Mumford-Shah Image Segmentation [J]. IEEE Transactions on Image Processing, 2006, 15(5): 1171-1181.
- [4] Mumford D, Shah J. Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems [J]. Communications on Pure and Applied Mathematics, 1989, 42(5): 577-685.
- [5] Tsai A, Yezzi A, Willsky A S. Curve Evolution Implementation of the Mumford-Shah Functional for Image Segmentation, Denoising, Interpolation, and Magnification [J]. IEEE Transactions on Image Processing, 2001, 10(8): 1169-1186.
- [6] 张少华, 何传江, 陈强. 结合全局和局部信息的活动轮廓模型 [J]. 计算机工程, 2011, 37(17): 203-205.
- [7] Ronfard R. Region-based Strategies for Active Contour Models [J]. International Journal of Computer Vision, 1994, 13(2): 223-247.
- [8] Zhu Guopu, Zhang Shunqun, Zeng Qingshuang, et al. Boundary-based Image Segmentation Using Binary Level Set Method [J]. Optical Engineering Letters, 2007, 46(5): 1-3.
- [9] Caselles V, Kimmel R, Sapiro G. Geodesic Active Contours [J]. International Journal of Computer Vision, 1997, 22(1): 61-79.
- [10] Chan T, Vese L. Active Contours Without Edges [J]. IEEE Transactions on Image Processing, 2001, 10(2): 266-277.
- [11] Li Chunming, Kao Chiuyuen, Gore J C, et al. Minimization of Region-scalable Fitting Energy for Image Segmentation [J]. IEEE Transactions on Image Processing, 2008, 17(10): 1940-1949.
- [12] Zhang Kaihua, Song Huihui, Zhang Lei. Active Contours Driven by Local Image Fitting Energy [J]. Pattern Recognition, 2010, 43(4): 1199-1206.

编辑 刘冰

(上接第 235 页)

- [5] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2005: 886-893.
- [6] 刘宏, 普杰信. 一种改进的自然场景特征提取方法 [J]. 计算机工程, 2011, 37(21): 182-184.
- [7] Bosch A, Zisserman A, Munoz X. Representing Shape with a Spatial Pyramid Kernel [C]//Proceedings of the 6th ACM International Conference on Image and Video. New York, USA: ACM Press, 2007: 401-408.
- [8] 赵碧茹. 场景特征可分性与协同求解方法 [D]. 合肥:

合肥工业大学, 2012.

- [9] 黄荣娟, 姜佳欣, 唐银凤, 等. 基于多特征组合的图像纹理分类 [J]. 计算机应用与软件, 2011, 28(8): 12-16.
- [10] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition [J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
- [11] Vapnik V. The Nature of Statistical Learning Theory [M]. 2nd ed. Berlin, Germany: Springer-Verlag, 2000.
- [12] Chang Chih-Chung, Lin Chih-Jen. LIBSVM: A Library for Support Vector Machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27.

编辑 刘冰