

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2013 3
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 3 月 第 18 卷 第 3 期(总第 203 期)

目次

图像处理和编码

沿边局部灰度差分椒盐噪声滤波 孔勇奇, 潘志庚(249)

图像分析和识别

核稀疏保留投影及在步态识别中的应用 王科俊, 阎涛, 吕卓纹, 唐墨(257)

局部 Gist 特征匹配核的场景分类 杨昭, 高隼, 谢昭, 吴克伟(264)

图像理解和计算机视觉

条件随机场模型的场景描述 赵龙, 郭立, 谢锦生(271)

基于 L_1 与空间点分类的 3 维场景重建 李静, 杨宜民, 付根平, 陈建平(277)

稀疏表示的 Lucas-Kanade 目标跟踪 徐如意, 陈靛影(283)

医学图像处理

管状特性和主动轮廓的 3 维血管自动提取 尚岩峰, 汪辉, 汪宁, 陈杰, 姚丽萍, 孙琨(290)

“第 17 届 CAD&CG 学术会议”专栏

多步攻击的规则树检测及可视化 胡亮, 赵剑明, 解男男, 努尔布力(299)

显著性边缘引导下的图像抽象化 王勋, 李红, 刘春晓, 严彩萍(305)

稀疏表达的运动数据压缩 齐天, 肖俊, 庄越挺(311)

基于语义的用户关注度计算 陈蕾英, 陈佳舟, 潘斌, 彭群生(318)

九宫格空间框架的图书图像检索 李宗民, 唐志辉(325)

光学相干层析视网膜体数据的 3 维分割 樊鲁杰, 孙延奎, 张田, 田小林(330)

混合软影绘制算法 刘浏, 周炜, 李华(336)

基于场的人群运动仿真 赵欣欣, 张勇, 孔德慧, 尹宝才(344)

利用大位移视图的自动可信图像修补 刘春晓, 金剑秋, 彭群生(351)

第八届国际数字地球会议征稿通知 (358)

Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 18 No. 3 March 2013

Contents

Image Processing and Coding

Salt and pepper noise filtering algorithm based on local border gray-scale differences Kong Yongqi, Pan Zhigeng(249)

Image Analysis and Recognition

Kernel sparsity preserving projections and its application to gait recognition Wang Kejun, Yan Tao, Lv Zhuowen, Tang Mo(257)

Scene categorization of local Gist feature match kernel Yang Zhao, Gao Jun, Xie Zhao, Wu Kewei(264)

Image Understanding and Computer Vision

Scene description based on the conditional random fields model Zhao Long, Guo Li, Xie Jinsheng(271)

Three dimensional scene reconstruction based on L_1 and scene point classification
..... Li Jing, Yang Yimin, Fu Genping Chen Jianping(277)

Lucas-Kanade tracking based on sparse representation Xu Ruyi, Chen Jingying(283)

Medical Image Processing

Three dimensional vessel extraction model based on tubular characters and active contour model
..... Shang Yanfeng, Wang Hui, Wang Ning, Chen Jie, Yao Liping, Sun Kun(290)

Issum of the CAD&CG'2012

Multi-step attacks detected by rules tree and visualization Hu Liang, Zhao Jianming, Xie Nannan, Nurbol(299)

Salient edge guided image abstraction optimization Wang Xun, Li Hong, Liu Chunxiao, Yan Caiping(305)

Motion data compression using sparse representation Qi Tian, Xiao Jun, Zhuang Yueting(311)

Semantic-based user interests computation Chen Leiying, Chen Jiazhou, Pan Bin, Peng Qunsheng(318)

Books image retrieval based on grid space frame Li Zongmin, Tang Zhihui(325)

Three dimensional segmentation to detect retinal boundary surfaces from OCT volume data
..... Fan Lujie, Sun Yankui, Zhang Tian, Tian Xiaolin(330)

Hybrid soft shadow rendering method Liu Liu, Zhou Wei, Li Hua(336)

Field-based crowd simulation Zhao Xinxin, Zhang Yong, Kong Dehui, Yin Baocai(344)

Automatic and accurate image completion from a large displacement view Liu Chunxiao, Jin Jianqiu, Peng Qunsheng(351)

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2013)03-0264-07

论文引用格式: 杨昭,高隽,谢昭,吴克伟. 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.

局部 Gist 特征匹配核的场景分类

杨昭, 高隽, 谢昭, 吴克伟

合肥工业大学计算机与信息学院, 合肥 230009

摘要: 针对场景分类任务中全局 Gist 特征粒度较为粗糙的问题, 提出一种基于稠密网格的局部 Gist 特征描述, 利用空间金字塔结构加入空间信息, 通过引入 RGB 颜色空间加入颜色信息, 并基于词汇包(BOW)模型设计一种高效匹配核来度量局部特征间的相似性, 核化特征匹配过程, 使用线性 SVM 完成场景分类。实验考察了不同尺度、方向、粒度和不同匹配核的局部 Gist 特征以及训练样本集的大小对分类结果的影响, 并通过在 OT 场景图像集上与全局 Gist 特征和稠密 SIFT 特征的场景分类结果进行比较, 充分说明了本文特征构造方法和分类模型的有效性。

关键词: 局部 Gist 特征; 空间金字塔; 高效匹配核; 场景分类

Scene categorization of local Gist feature match kernel

Yang Zhao, Gao Jun, Xie Zhao, Wu Kewei

School of Computer and Information, Hefei University of Technology, Hefei 230009, China

Abstract: Due to the coarse fineness of global Gist features in scene categorization tasks, we propose a local Gist feature description based on a dense grid. It uses a spatial pyramid structure to add distribution information and introduces the RGB color space to add color information. The feature matching process is kernelized by an efficient match kernel which measures the similarity between local features based on the BOW model. The scene categorization task can be done with linear SVM. Experiment shows the influence to the classification accuracy with local Gist features which have different scale, orientation, fineness, match kernels and numbers of training samples. By using the classification result of the global Gist feature and dense SIFT features on the OT scene dataset, we demonstrate that the proposed feature construction method and classification model are efficient.

Key words: local Gist feature; spatial pyramid; efficient match kernel; scene classification

0 引言

场景分类是计算机视觉和图像理解领域的一个重要研究方向,其本质是对场景图像进行某种量化分析,由给定的一组语义类别对图像数据库进行自动标注。

基于图像中层语义建模的场景分类是目前常用且有效的一类方法,主要包括语义对象分割法、

局部语义概念法和全局语义建模法,其中后两种方法无需对场景内容进行分割和标注。局部语义概念法使用关键点周围的局部描述子对场景建模来完成分类任务,最有代表性的是词汇包^[1](BOW)模型。BOW模型将每个局部特征用最接近的视觉词汇表示,视觉词汇形成视觉词典或码本,再根据其统计直方图表征场景,用直方图交集等方法度量场景图像间的相似性。生成视觉词汇的方法有很多种^[2],最常用的是 k-means 聚类。针

收稿日期:2012-07-12;修回日期:2012-09-11

基金项目:国家自然科学基金项目(60905005,6875012,61273237);教育部博士点基金项目(20090111110015)

第一作者简介:杨昭(1989—),男,2010年合肥工业大学在读电子信息工程硕士研究生,主要研究方向为图像理解、人工智能、模式识别。E-mail: yangzhao_ez@163.com

对视觉词汇的无序性,Lazebnik^[3]等人提出金字塔匹配(SPM)模型,利用视觉词汇的空间布局信息实现场景的有效分类。全局语义建模法将场景作为一个整体进行全局描述,利用全局特征完成场景分类,如Oliva和Torralba^[4-5]通过Gabor滤波器提取图像不同频率和方向上的信息,再将滤波后的图像划分为 4×4 规则网格,对每个网格取均值,级联形成全局Gist特征,对场景的整体结构和布局进行描述,但这种稀疏网格划分下的Gist特征是粗粒度的,因为同一网格中可能包含不同的场景部分,取均值后会丢失场景细节信息。

近年来核方法^[6](KM)是机器学习领域的一个新热点,对一些只涉及样本间内积运算的机器学习算法,如果改变内积定义的方式,利用核技巧,即用事先定义的核函数取代内积,就可以得到与该学习方法对应的核方法。核方法相比其他机器学习方法,如神经网络,由于其扎实的数学基础而受到青睐^[7],其最成功的应用是支持向量机^[8](SVM)理论。核函数的采用使得线性SVM很容易地推广到非线性SVM,其核心在于利用相对简单的核函数运算,既避免了特征空间中复杂的内积计算,又避免了特征空间(学习机器)本身的设计^[9]。

针对全局Gist特征粗粒度的问题,提出一种基于稠密网格的局部Gist特征描述,由局部Gist特征聚类形成视觉词典,使用匹配核来完成场景图像的矢量化过程,并加入了空间和颜色信息。由于匹配核的引入相当于将SVM中核函数的作用提前,所以可以使用线性分类器,而不必显式地计算整个核矩阵,对图像数和特征数来说计算代价都是线性的。实验选用OT场景图像集,考察了不同尺度、方向、粒度和不同匹配核的局部Gist特征以及训练样本集的大小对分类结果的影响,并同基于全局Gist特征和稠密SIFT特征的场景分类结果进行了比较,充分说明了本文方法的有效性。

1 场景 Gist 特征提取

Gist特征^[4-5]是一种生物启发特征,该特征模拟人的视觉提取图像中粗略但简明扼要的上下文信息。场景Gist特征是通过多尺度多方向Gabor滤波器组对场景图像滤波后得到的轮廓信息。

1.1 Gabor 滤波器组

在信号处理中1维Gabor函数被广泛使用,

1985年Daugman^[10]将其扩展到2维形式,利用2维Gabor函数可以形成Gabor滤波器,表达式为

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \times \cos(2\pi f_0 x + \varphi) \quad (1)$$

式中, x 和 y 是图像像素坐标位置, σ_x 和 σ_y 分别是 x 和 y 方向Gaussian因子的方差, f_0 是滤波器中心频率,参数 φ 是该谐波因子的相位差。

多尺度多方向Gabor滤波器组是在Gabor滤波器 $g(x, y)$ 基础上通过尺度和旋转变换扩展而成的多通道滤波方案,是一种自相似Gabor小波,即

$$\begin{aligned} g_{mn}(x, y) &= a^{-m} g(x', y'), a > 1 \\ x' &= a^{-m} (x \cos \theta + y \sin \theta) \\ y' &= a^{-m} (-x \sin \theta + y \cos \theta) \\ \theta &= n\pi / (n + 1) \end{aligned} \quad (2)$$

式中, a^{-m} 为母小波膨胀的尺度因子, θ 为旋转角度,即滤波器的方向。 m 为Gabor滤波器组的尺度数, n 为Gabor滤波器组的方向数,改变 m, n 即可得到一组不同的Gabor小波滤波器。

1.2 图像 Gist 特征提取

利用Gabor滤波器对图像进行特征提取的实质就是将一组Gabor小波函数基与图像进行卷积操作。这组Gabor小波函数基中的每一个Gabor函数都会在与它的振荡方向相垂直的边缘处产生强烈的响应,因此能够检测出图像中的一些具有相应的方向频率信息的显著特征,从而形成了对输入的原图像的一种鲁棒的、紧凑的特征表示。

一幅大小为 $r \times c$ 的灰度图像 $f(x, y)$ 用 m 尺度 n 方向的Gabor滤波器组进行滤波,即分别同 n_c 个通道的滤波器进行卷积,其中 $n_c = m \times n$,再级联卷积的结果,称为图像Gist(IG)特征,即

$$G'(x, y) = \underset{n_c}{cat}(f(x, y) * g_{mn}(x, y)) \quad (3)$$

式中, cat 为级联运算符, $*$ 为卷积运算符, G' 的维数为 $n_c \times r \times c$ 。

利用4尺度8方向的Gabor滤波器组进行图像Gist特征提取的结果如图1所示。

图1(b)就是Gabor滤波器组中32个滤波器分别对图像进行滤波后级联的结果,可见该场景内的主要结构部分已被提取出来。

1.3 全局 Gist 特征提取

将一幅大小为 $r \times c$ 的灰度图像 $f(x, y)$ 划分成 $n_p \times n_p$ 的规则网格,则网格块数为 $n_g = n_p \times n_p$ 。各

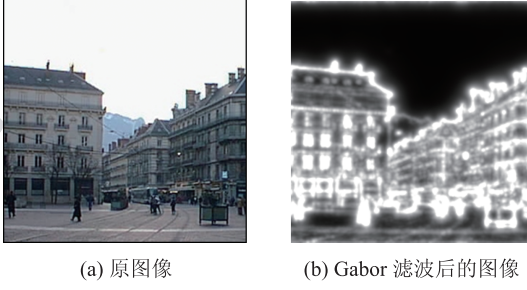


图 1 Gabor 滤波器组提取的图像 Gist 特征
Fig. 1 IG feature extracted by Gabor filter group

网格块按行依次记作 P_i , 其中 $i = 1, \dots, n_g$ 。网格块大小为 $r' \times c'$, 其中 $r' = r/n_p, c' = c/n_p$ 。

分别用 n_c 个通道的滤波器对图像进行卷积滤波, 则每个网格块各通道滤波后级联的结果称为块 Gist (PG) 特征, 即

$$G_i^P(x, y) = \text{cat}_{n_c}(f(x, y) * g_{mn}(x, y)) \quad (4)$$

$$(x, y) \in P_i$$

式中, G^P 的维数为 $n_c \times r' \times c'$, 且 G^P 是 G^I 的子集, 即 $G^P \subseteq G^I$ 。

对 G^P 各通道滤波结果取均值后按行组合的结果称为全局 Gist (GG) 特征, 即

$$G^G = \{\overline{G_1^P}, \overline{G_2^P}, \dots, \overline{G_{n_g}^P}\} \quad (5)$$

式中 $\overline{G_i^P} = \frac{1}{r' \times c'} \sum_{(x,y) \in P_i} G_i^P(x, y)$, G^G 的维数为 $n_c \times n_g$ 。

Oliva 和 Torralba^[4-5] 将场景图像划分为 4×4 的规则网格, 用 4 尺度 8 方向共 32 个 Gabor 滤波器得到全局 Gist 特征, 其提取过程如图 2 所示。

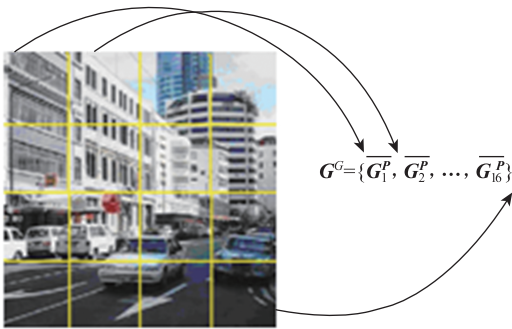


图 2 全局 Gist 特征提取过程
Fig. 2 GG feature extracting process

这种全局 Gist 特征虽然保持了场景图像的全局结构和布局, 但是 4×4 的网格划分不够精细, 单个网格内的场景可能会包含不同的部分, 如图 2 的网格内会同时有建筑和天空部分等。

1.4 局部 Gist 特征提取

为了得到细粒度的 Gist 特征, 采用 8 像素间隔的稠密网格划分, 网格大小固定为 16×16 。先计算每个网格块的 G^P 。然后将网格划分成 $n_b \times n_b$ 的子区域, 则子区域数为 $n_z = n_b \times n_b$ 。各子区域按行依次记作 Z_j , 其中 $j = 1, \dots, n_z$ 。各子区域中所有像素的 G^P 值加权求和, 权重 w 为对应像素到子区域中心像素归一化距离的倒数, 得到的结果称为区域 Gist (ZG) 特征, 即

$$G_j^Z(x, y) = w_j(x, y) G_j^P(x, y) \quad (6)$$

$$(x, y) \in Z_j$$

式中, w_j 为各子区域中像素对应的权重, G^Z 的维数为 $n_c \times 1$ 。

将各网格块的 G^Z 按行组合的结果称为局部 Gist (LG) 特征, 即

$$G^L = \{G_1^Z, G_2^Z, \dots, G_{n_z}^Z\} \quad (7)$$

G^L 的维数为 $n_c \times n_z$ 。稠密网格的划分使场景中不同的部分基本得以区分, 而子区域的划分进一步细化了 Gist 特征的粒度。

对于场景图像来说颜色信息十分重要, 所以引入 RGB 颜色空间, 分别对每个网格块在 RGB 3 个颜色通道上计算局部 Gist 特征, 级联得到的结果称为颜色局部 Gist (CLG) 特征, 即

$$G_c^L = \text{cat}(G_r^L, G_g^L, G_b^L) \quad (8)$$

式中, G_r^L, G_g^L, G_b^L 分别为 RGB 颜色通道上的局部 Gist 特征, G_c^L 的维数为 $n_c \times n_z \times 3$ 。

2 局部特征匹配

在有监督图像分类任务中, 已知训练图像集和对应的标签, 目标是学习一个分类器来得到测试图像集的标签。如果直接使用局部 Gist 特征, 则维数太高, 所以引入 BOW 模型, 形成视觉词汇和词典, 提出一种高效的匹配核, 核化场景分类过程, 并用空间金字塔结构加入空间布局信息。

2.1 词汇包模型

两幅图像 A 和 B 的 LG 特征集合分别记作 $F^A = \{G_1^{LA}, \dots, G_p^{LA}\}, F^B = \{G_1^{LB}, \dots, G_p^{LB}\}$, 其中 p 为图像划分的网格数, $V = \{v_1, \dots, v_D\}$ 为视觉词汇集合, 即视觉词典。

在词汇包 (BOW) 模型中, 每个局部特征量化成一个 D 维的二值示性向量

$$\mu(\mathbf{G}^L) = [\mu_1(\mathbf{G}^L), \dots, \mu_D(\mathbf{G}^L)]^T \quad (9)$$

式中, $\mu_i(\mathbf{G}^L) = \begin{cases} 1 & \mathbf{G}^L \in R(\mathbf{v}_i) \\ 0 & \text{其他} \end{cases}$, $\exists i = 1, \dots, D$,
 $R(\mathbf{v}_i) = \{\mathbf{G}^L: \|\mathbf{G}^L - \mathbf{v}_i\| \leq \|\mathbf{G}^L - \mathbf{v}\|, \forall \mathbf{v} \in V\}$ 。

由归一化直方图得到一幅图像的特征向量

$$\bar{\mu}(\mathbf{F}) = \frac{1}{|\mathbf{F}|} \sum_{\mathbf{G}^L \in \mathbf{F}} \mu(\mathbf{G}^L), \text{ 式中 } |\cdot| \text{ 表示集合的势。}$$

传统 BOW 模型用视觉词汇的统计直方图来表征场景, 然后对得到的 BOW 特征使用非线性分类器, 如 SVM 核分类器, 但这往往会需要很高的训练和测试代价。

2.2 高效匹配核

为了能够使用线性分类器, 引入匹配核

$$K(\mathbf{F}^A, \mathbf{F}^B) = \bar{\mu}(\mathbf{F}^A)^T \bar{\mu}(\mathbf{F}^B) = \frac{1}{|\mathbf{F}^A| |\mathbf{F}^B|} \sum_{\mathbf{G}^{LA} \in \mathbf{F}^A} \sum_{\mathbf{G}^{LB} \in \mathbf{F}^B} \mu(\mathbf{G}^{LA})^T \mu(\mathbf{G}^{LB}) = \frac{1}{|\mathbf{F}^A| |\mathbf{F}^B|} \sum_{\mathbf{G}^{LA} \in \mathbf{F}^A} \sum_{\mathbf{G}^{LB} \in \mathbf{F}^B} \delta(\mathbf{G}^{LA}, \mathbf{G}^{LB}) \quad (10)$$

$$\text{式中, } \delta(\mathbf{G}^{LA}, \mathbf{G}^{LB}) = \begin{cases} 1 & \mathbf{G}^{LA}, \mathbf{G}^{LB} \subset R(\mathbf{v}_i) \\ 0 & \text{其他} \end{cases}。$$

匹配核将非线性 SVM 中核函数的作用提前, 完成了图像特征向量的矢量量化过程, 在特征向量中选出与视觉词典距离最近的局部特征来表示, 从而降低了特征维数。

$\delta(\mathbf{G}^{LA}, \mathbf{G}^{LB})$ 是一个正定核, 用来度量图像 LG 特征 $\mathbf{G}^{LA}$ 和 $\mathbf{G}^{LB}$ 间的相似性: $\mathbf{G}^{LA}$ 和 $\mathbf{G}^{LB}$ 属于同一区域 $R(\mathbf{v}_i)$ 时, $\delta(\mathbf{G}^{LA}, \mathbf{G}^{LB}) = 1$, 但是这种量化标准太过粗糙, 会导致分类效果下降, 所以一般用连续核函数 $k(\mathbf{G}^{LA}, \mathbf{G}^{LB})$ 来替代, 即

$$K_s(\mathbf{F}^A, \mathbf{F}^B) = \frac{1}{|\mathbf{F}^A| |\mathbf{F}^B|} \sum_{\mathbf{G}^{LA} \in \mathbf{F}^A} \sum_{\mathbf{G}^{LB} \in \mathbf{F}^B} k(\mathbf{G}^{LA}, \mathbf{G}^{LB}) \quad (11)$$

式中 $k(\mathbf{G}^{LA}, \mathbf{G}^{LB}) = \phi(\mathbf{G}^{LA})^T \phi(\mathbf{G}^{LB})$, 称为局部匹配核。如果局部特征映射 $\phi(\mathbf{G}^L)$ 是有限维的, 则由式(11)简化可得到高效匹配核

$$K_s(\mathbf{F}^A, \mathbf{F}^B) = \bar{\phi}(\mathbf{F}^A)^T \bar{\phi}(\mathbf{F}^B) \quad (12)$$

式中, $\bar{\phi}(\mathbf{F}) = \frac{1}{|\mathbf{F}|} \sum_{\mathbf{G}^L \in \mathbf{F}} \phi(\mathbf{G}^L)$ 。高效匹配核只需计算特征向量 $\bar{\phi}(\mathbf{F})$, 而不用计算整个核矩阵。高效匹配核的训练代价为 $O(nmDd)$, 每幅图像的测试代价为 $O(mDd)$, 其中 n 为图像数, m 为特征集合的平均势, d 为局部特征 $\mathbf{G}^L$ 的维数, D 为局部特征映射 $\phi(\mathbf{G}^L)$ 的维数。如果局部特征映射 $\phi(\mathbf{G}^L)$ 是低维的, 则高效匹配核比式(11)匹配核计算代价低得

多, 式(11)匹配核的训练代价和测试代价分别为 $O(n^2 m^2 d)$ 和 $O(nm^2 d)$ 。例如, 当 D 和 m 等阶 (本文实验的情况) 时, 高效匹配核的计算代价仅为式(11)匹配核的 $1/n$ 。

匹配核函数有很多种选择, 比较常用的是线性 (linear) 函数 $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ 、拉普拉斯 (laplace) 函数 $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-(\alpha \mathbf{x}_i + \beta \mathbf{x}_j))$ 和径向基 (rbf) 函数 $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \eta)$, 其中样本 $\mathbf{x}_i, \mathbf{x}_j$ 均为列向量, α, β, η 分别为核函数中的参数, 常取为常数。

空间信息可以通过空间金字塔结构^[3] 加到高效匹配核中, 即

$$K_p(\mathbf{F}^A, \mathbf{F}^B) = \sum_{l=0}^{L-1} \sum_{t=1}^{2^l} 2^{-l} K_s(\mathbf{F}_{(l,t)}^A, \mathbf{F}_{(l,t)}^B) = \bar{\phi}_p(\mathbf{F}^A)^T \bar{\phi}_p(\mathbf{F}^B) \quad (13)$$

式中, L 是金字塔层数, 2^l 是第 l 层中空间单元数, $\mathbf{F}_{(l,t)}^A$ 是落入空间单元 (l, t) 中的局部特征, $\bar{\phi}_p(\mathbf{F}^A) = [\bar{\phi}(f_{(1,1)}^A)^T, \dots, \bar{\phi}(f_{(L,2^L)}^A)^T]^T$ 。

2.3 低维局部特征映射求解

令 $k(\mathbf{G}^{LA}, \mathbf{G}^{LB}) = \psi(\mathbf{G}^{LA})^T \psi(\mathbf{G}^{LB})$, 其中 $\psi(\mathbf{G}^L)$ 表示高维局部特征映射, 对 $\psi(\mathbf{G}^L)$ 进行低维投影来得到高效匹配核对应的低维局部特征映射 $\phi(\mathbf{G}^L)$, 使其内积近似于给定的匹配核函数。

先对词典核矩阵 $\mathbf{K}_V = \mathbf{K}_s(\mathbf{V}, \mathbf{V})$ 做奇异值分解 (SVD), 得到基向量映射集 $\{\psi(\mathbf{z}_i)\}_{i=1}^D$, 则 $\psi(\mathbf{G}^{LA})$ 可近似求解

$$\bar{\omega}_A = \arg \min_{\omega_A} \|\psi(\mathbf{G}^{LA}) - \mathbf{H} \omega_A\|^2 \quad (14)$$

式中, $\mathbf{H} = [\psi(\mathbf{z}_1), \dots, \psi(\mathbf{z}_D)]$, $\bar{\omega}_A$ 是低维投影系数。式(14)是一个凸二次规划问题, 有解析解

$$\bar{\omega}_A = (\mathbf{H}^T \mathbf{H})^{-1} (\mathbf{H}^T \psi(\mathbf{G}^{LA})) \quad (15)$$

由投影向量得到的局部匹配核为

$$k_l(\mathbf{G}^{LA}, \mathbf{G}^{LB}) = [\mathbf{H} \bar{\omega}_A]^T [\mathbf{H} \bar{\omega}_B] = \mathbf{k}_Z(\mathbf{G}^{LA})^T \mathbf{K}_{ZZ}^{-1} \mathbf{k}_Z(\mathbf{G}^{LB}) \quad (16)$$

式中, $\mathbf{k}_Z$ 是一个 $D \times 1$ 维向量: $\{\mathbf{k}_Z\}_i = k(\mathbf{G}^{LA}, \mathbf{z}_i)$, $\mathbf{K}_{ZZ}$ 是一个 $D \times D$ 维矩阵: $\{\mathbf{K}_{ZZ}\}_{ij} = k(\mathbf{z}_i, \mathbf{z}_j)$ 。

令 $\mathbf{G}^T \mathbf{G} = \mathbf{K}_{ZZ}^{-1}$, 则低维局部特征映射为

$$\phi(\mathbf{G}^{LA}) = \mathbf{G} \mathbf{k}_Z(\mathbf{G}^{LA}) \quad (17)$$

最后得到高效匹配核对应的特征映射为

$$\bar{\phi}(\mathbf{F}^A) = \frac{1}{|\mathbf{F}^A|} \mathbf{G} \left[\sum_{\mathbf{G}^{LA} \in \mathbf{F}^A} \mathbf{k}_Z(\mathbf{G}^{LA}) \right] \quad (18)$$

2.4 局部特征匹配算法

算法步骤如下:

1) 局部特征计算。按式(7)计算输入图像的 LG 特征。

2) 低维特征映射计算:

(1) 按式(15)计算低维投影系数 $\bar{\omega}$;

(2) 按式(16)计算局部匹配核和 K_{zz}^{-1} ;

(3) 按式(18)计算高效匹配核对应的特征映射 $\bar{\phi}(F)$ 。

3) 高效匹配核计算, 按式(13)计算加入空间信息的高效匹配核 $K_p(F^A, F^B)$ 。

测试图像用高效匹配核进行局部特征匹配,

完成矢量量化过程后, 可使用线性 SVM 实现分类。

3 实验分析

选用图像数据集为 MIT 的 Oliva&Torralba (OT) 室外场景图像集, 包含如图 3 所示的海滨、森林、高速公路、城市、高山、乡村、街道和高楼 8 个类别, 4 类自然场景图像, 4 类人造场景图像, 每幅大小为 256×256 , 共 2688 幅。



图 3 OT 图像数据集包含的 8 个类别室外场景图像

Fig. 3 Eight classes of outdoor scene images included in OT dataset

3.1 场景分类及相关参数讨论

先不考虑颜色信息和 Gabor 滤波器不同尺度数和方向数的影响, 采用 8 像素间隔的稠密网格来计算场景图像局部 Gist 特征, 网格大小为 16×16 , 用 K 均值聚类得到视觉词汇, 特征映射维数和词典大小为 1000。匹配核函数分别选择: 线性 (linear) 函数、拉普拉斯 (laplace) 函数和径向基 (rbf) 函数, 并加入 3 层金字塔空间结构。

分类器选用线性 SVM, 用一对多 (1vsRest) 方式拓展成多类判决。SVM 的正则项和核参数由交叉验证来调整。每类训练样本数分别取 50、60、70、80、90 幅时进行实验, 训练集外其他所有图像用做测试, 用 5 次随机训练和测试的分类平均正确率来评价结果, 如图 4 所示。可见训练集样本个数从 50 增至 90, linear 和 laplace 匹配核的分类正确率基本不变, 而 rbf 匹配核的分类正确率略有提高, 所以后面的实验每类训练样本数固定为 50。

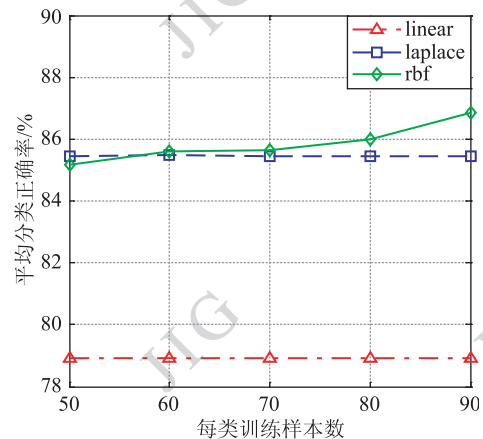


图 4 不同匹配核和训练样本数下的平均分类正确率

Fig. 4 Classification accuracy on different match kernels and numbers of training samples

接着讨论不同尺度、方向、粒度和不同匹配核的局部 Gist 特征对分类结果的影响。局部 Gist 特征计算过程中若每个网格划分成 $n_b \times n_b$ 个子区域, 则

称该特征的粒度为 n_b 。当 $n_b = 1$ 时,整个网格相当于一个子区域,为最粗粒度;当 $n_b = 16$ 时,网格内每个像素相当于一个子区域,为最细粒度。下面的实验中局部 Gist 特征分别取粒度 $n_b = 1, \dots, 4$,尺度数为 2、4、6,方向数为 4、8、12,匹配核函数为 linear、laplace 和 rbf 函数,实验结果如表 1 所示。

从表 1 的数据可以看出,匹配核函数选择 rbf 函数,Gabor 滤波器选择 4 尺度 8 方向,粒度选择 4×4 子区域划分的局部 Gist 特征分类正确率最高,但粒度也不是越细越好, 4×4 的子区域划分在有些情况下反而比 3×3 的子区域划分效果差,所以粒度 4 为最佳选择。

表 1 不同尺度、方向、粒度和匹配核的局部 Gist 特征平均分类正确率

Table 1 Classification accuracy of LG feature with different scales, orientations, finenesses and match kernels

	尺度数	方向数 4 粒度				方向数 8 粒度				方向数 12 粒度			
		1	2	3	4	1	2	3	4	1	2	3	4
linear	2	69.43	62.69	76.96	77.40	74.03	61.73	79.40	78.80	60.13	73.94	79.86	78.19
	4	72.24	59.80	80.36	79.60	78.91	69.64	81.78	82.38	69.53	73.49	82.97	81.99
	6	67.09	69.51	81.52	80.37	76.97	72.36	80.96	83.04	68.58	77.27	81.76	80.23
laplace	2	83.15	81.74	82.20	81.81	83.21	80.66	82.28	81.30	80.14	82.14	81.60	80.12
	4	84.36	82.21	82.88	82.86	85.44	83.23	84.24	82.60	82.45	83.12	84.04	77.14
	6	82.03	81.93	82.66	81.30	84.24	83.87	83.30	75.60	78.60	83.50	80.53	64.86
rbf	2	80.32	81.09	84.36	84.86	83.40	79.94	85.53	84.95	80.02	83.27	85.33	84.93
	4	82.50	79.68	85.94	85.78	85.63	84.50	86.22	86.85	80.64	83.92	86.45	86.14
	6	79.63	81.30	86.07	86.15	83.30	82.98	85.79	85.72	77.39	83.18	85.19	84.81

3.2 与其他方法对比

Oliva 和 Torralba^[4-5]提出的全局 Gist 特征使用的是灰度图像,并没有包含颜色信息,而颜色信息对于场景图像分类十分重要。本文提出的局部 Gist 特征可以通过引入 RGB 颜色空间,拓展成颜色局部 Gist 特征。SIFT^[11]是一种基于梯度的图像块描述子,是目前最有效、最具代表性的局部特征。采用网格大小为 16×16 、网格块间隔为 8 的稠密网格来计算 SIFT 特征,便于和本文的局部 Gist 特征比较,进一步验证本文方法的优越性。

基于 SIFT 特征、全局 Gist (GG) 特征、局部 Gist (LG) 特征和颜色局部 Gist (CLG) 特征的场景分类实验结果如表 2 所示,每类场景训练样本数都为 50,其中 Gist 特征都选用 4 尺度 8 方向的 Gabor 滤波器,而 SIFT 特征、局部 Gist 特征和颜色局部 Gist 特征选择 4×4 的子区域划分,匹配核函数选择 rbf 函数,并加入 3 层金字塔空间结构。分类器选用线性 SVM,用一对多方式拓展成多类判决。SVM 的正则项和核参数由交叉验证来调整,用 5 次随机训练和测试的分类平均正确率来评价结果。

表 2 SIFT、全局 Gist、局部 Gist 和颜色局部 Gist 特征对 OT 场景数据集分类性能比较

Table 2 Comparison of the performance with SIFT, GG, LG and CLG on OT scene dataset

场景	SIFT	GG	LG	CLG
高楼	92.48	88.08	92.16	92.55
城市	87.60	83.37	89.92	90.22
街道	88.43	87.42	90.08	91.07
高速公路	82.38	85.17	82.86	84.43
海滨	80.00	81.47	85.16	86.32
乡村	77.78	70.69	75.83	77.80
高山	83.95	79.22	84.88	87.24
森林	97.12	92.99	93.88	97.17
平均	86.22	83.55	86.85	88.35

从表 2 的实验结果可以看出,本文的局部 Gist 特征取得了较好的分类结果,在同一分类模型框架下优于全局 Gist 特征和 SIFT 特征的效果。8 类场景的平均分类正确率在 86% 以上,而增加了颜色信息后的颜色局部 Gist 特征使得分类正确率进一步

提高到88%以上,特别是对较难识别的城市、海滨和高山场景识别率有明显提高。

4 结 论

针对场景分类任务中全局 Gist 特征粗粒度的问题,提出了一种基于稠密网格的局部 Gist 特征描述,加入了空间和颜色信息,结合 BOW 模型引入高效匹配核,使用线性分类器。实验说明使用 rbf 匹配核、4 尺度 8 方向、粒度为 4 的局部 Gist 特征分类正确率最高,特别是颜色局部 Gist 特征,全面优于全局 Gist 特征和稠密 SIFT 特征的场景分类结果。讨论新的网格、子区域划分方法和高效匹配核计算过程中低维局部特征映射的求解方法,以及在其他场景图像集上验证局部 Gist 特征的有效性,是未来工作的研究重点。

参考文献 (References)

- [1] Sivic J, Zisserman A. Video google: a text retrieval approach to object matching in videos [C] // Proceedings of International Conference on Computer Vision. Washington DC: [s. n.], 2003:1470-1477.
- [2] Jurie F, Triggs B. Creating efficient codebooks for visual recognition [C] // Proceedings of International Conference on Computer Vision. Beijing: [s. n.], 2005: 604-610.
- [3] Lazebnik S, Schmid C. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2006, 2:2169-2178.
- [4] Oliva A, Torralba A. Modeling the shape of the scene a holistic representation of the spatial envelope [J]. International Journal in Computer Vision, 2001, 42(3): 145-175.
- [5] 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.
- [6] Muller K R, Mika S, Ratsch G, et al. An introduction to kernel based learning algorithms [J]. IEEE Transactions on Neural Networks, 2001, 12(2): 181-201.
- [7] Hofman T, Schölkopf B. Kernel methods in machine learning [J]. The Annals of Statistics, 2008, 36(3): 1171-1220.
- [8] Vapnik V N. Statistical Learning Theory [M]. New York: Wiley, 1998.
- [9] Schölkopf B, Smola A J. Learning with Kernels [M]. Massachusetts: The MIT Press, 2002.
- [10] Daugman J. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters [J]. Journal of the Optical Society of America, 1985, 2(7): 1160-1169.
- [11] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2):91-110.