



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2013
08
VOL.18

ISSN1006-8961
CN11-3758/TB



北魏



隋



初唐

数字文化遗产研究 P968

综述

尺度不变特征变换算子综述

刘立, 詹茵茵, 罗扬, 刘朝晖, 彭复员 885

图像处理和编码

改进的H.264帧间模式选择算法

任克强, 张旭光, 罗会兰 893

均质雾天下摄像机动态标定的车速估计

宋洪军, 陈阳舟, 郜园园 901

基于人脸五官特征的空域差错掩盖算法

张江鑫, 谢晋, 邝万坤 913

采用圆谐-傅里叶矩的图像区域复制粘贴篡改检测

秦娟, 李峰, 向凌云, 殷蓓蓓 919

图像分析和识别

基于视频感知哈希的视频篡改检测与多粒度定位

朱映映, 文振焜, 杜以华, 邓良太 924

视觉注意模型的道路监控视频关键帧提取

刘云鹏, 张三元, 王仁芳, 张引 933

格拉斯曼流形上的半监督判别分析

姜伟, 陆瑶, 杨炳儒 944

精神疲劳实时监测中多面部特征时序分类模型

陈云华, 张灵, 丁伍洋, 严明玉 953

图像分类中的概率乘积核函数

杨赛, 赵春霞 961

轮廓整体结构约束的壁画图像相似性度量

唐大伟, 鲁东明, 杨冰, 许端清 968

应用色彩空间聚类方法实现道路建模

向宸薇, 王拓, 于舰 976

图像理解和计算机视觉

杂乱背景和摄像机移动下的时空兴趣点检测

刘长红, 陈勇, 王明文 982

形状和颜色混合不变量在图像检索中的应用

公明, 曹伟国, 李华 990

计算机图形学

结合统计分类和边缘特征的最优视图提取

潘翔, 陈敖, 章国栋 1004

医学图像处理

局部熵驱动下的脑MR图像分割与偏移场恢复耦合模型

张建伟, 杨红, 陈允杰, 方林, 詹天明 1011

改进的模糊聚类实质肺结节3维分割

李翠芳, 聂生东, 王远军, 孙希文, 郑斌 1019

遥感图像处理

高光谱海岸带区域分割的活动轮廓模型

王相海, 金弋博 1031

改进的基于非局部均值的极化SAR相干斑抑制

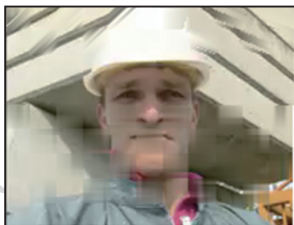
赵忠民, 赵拥军, 牛朝阳 1038

小波变换和稀疏表示相结合的遥感图像融合

刘婷, 程建 1045

《中国图象图形学报》论文摘要写作要求

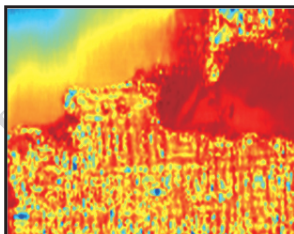
1054



基于人脸五官特征的空域差错掩盖算法 (第913页)



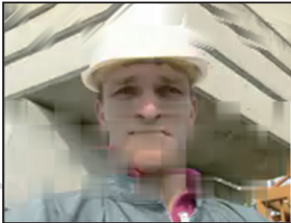
应用色彩空间聚类方法实现道路建模 (第976页)



改进的基于非局部均值的极化SAR相干斑抑制 (第1038页)

CONTENTS

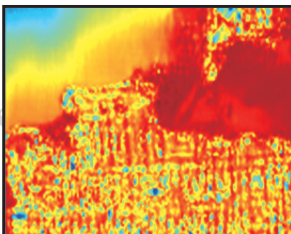
JOURNAL OF IMAGE AND GRAPHICS



Spatial error concealment algorithm based on the facial features (P913)



Road modeling using color spatial clustering (P976)



Improved polarimetric SAR speckle filter based on non-local means (P1038)

Review

- Summarization of the scale invariant feature transform
Liu Li , Zhan Yinyin, Luo Yang, Liu Chaohui, Peng Fuyuan 885

Image Processing and Coding

- Improved inter frame mode selection algorithm for H.264
Ren Keqiang, Zhang Xuguang, Luo Huilan 893
- Vehicle velocity estimation based on camera calibration under homogenous fog
Song Hongjun, Chen Yangzhou, Gao Yuanyuan 901
- Spatial error concealment algorithm based on the facial features
Zhang Jiangxin, Xie Jin, Kuang Wankun 913
- Detection of image region copy-move forgery using radial harmonic Fourier moments
Qin Juan, Li Feng, Xiang Lingyun, Yin Changming 919

Image Analysis and Recognition

- Video forgery detection and multi-granularity location based on video perceptual hashing
Zhu Yingying, Wen Zhenkun, Du Yihua, Deng Liangtai 924
- Key frame extraction based on the visual attention model for lane surveillance video
Liu Yunpeng, Zhang Sanyuan, Wang Renfang, Zhang Yin 933
- Semi-supervised discriminant analysis on Grassmannian manifold
Jiang Wei, Lu Yao, Yang Bingru 944
- Time-series classification model based on multiple facial feature for real-time mental fatigue monitoring
Chen Yunhua, Zhang Ling, Ding Wuyang, Yan Mingyu 953
- Probability product kernels in image classification
Yang Sai, Zhao Chunxia 961
- Similarity metrics between mural images with constraints of the overall structure of contours
Tang Dawei, Lu Dongming, Yang Bing, Xu Duanqing 968
- Road modeling using color spatial clustering
Xiang Chenwei, Wang Tuo, Yu Jian 976

Image Understanding and Computer Vision

- Spatio-temporal interest point detection in cluttered backgrounds with camera movements
Liu Changhong, Chen Yong, Wang Mingwen 982
- Application of combined shape and color invariants to image retrieval
Gong Ming, Cao Weiguo, Li Hua 990

Computer Graphics

- Best view selection based on statistical classification and edge features
Pan Xiang, Chen Ao, Zhang Guodong 1004

Medical Image Processing

- Brain MR image segmentation and bias correction model based on local entropy
Zhang Jianwei, Yang Hong, Chen Yunjie, Fang Ling, Zhan Tianming 1011
- Segmentation of sub-solid pulmonary nodules based on improved fuzzy C-means clustering
Li Cuifang, Nie Shengdong, Wang Yuanjun, Sun Xiwen, Zheng Bin 1019

Remote Sensing Image Processing

- The active contour model for segmentation of coastal hyperspectral remote sensing image
Wang Xianghai, Yin Yibo 1031
- Improved polarimetric SAR speckle filter based on non-local means
Zhao Zhongmin, Zhao Yongjun, Niu Chaoyang 1038
- Remote sensing image fusion with wavelet transform and sparse representation
Liu Ting, Cheng Jian 1045

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2013)08-0961-07

论文引用格式: 杨赛, 赵春霞. 图像分类中的概率乘积核函数[J]. 中国图象图形学报, 2013, 18(8): 961-967. [DOI: 10.11843/jig.20130810]

图像分类中的概率乘积核函数

杨赛, 赵春霞

南京理工大学计算机科学与技术学院, 南京 210094

摘要: Bag-of-Features (BOF) 模型使用编码矢量的某一特定统计值表征图像, 与基于传统核函数的支持向量机相配合完成对图像的分类, 所带来的问题是会丢失大量判别信息以及最优核函数的选择。因此, 本文将硬分配编码矢量服从的多项分布、软分配编码矢量服从 Dirichlet 分布, 并以此作为图像的内容描述, 利用最大似然算法估计其中参数, 然后使用概率乘积核函数计算图像两两之间的核函数, 最后使用支持向量机对图像进行分类。公开图像数据集上的实验结果表明, 本文算法取得了更优的分类性能。

关键词: BOF 模型; 统计值; 概率乘积核函数; 支持向量机; 图像分类

Probability product kernels in image classification

Yang Sai, Zhao Chunxia

School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract: Images are characterized by some statistics of coded vectors in the Bag-of-Features (BOF) model, and then classified by support vector machine (SVM) based on traditional kernel, the existing problems are the loss of discriminant information and choosing of optimal kernel. To solve these problems, we use the multinomial distribution of hard coded vectors or Dirichlet distribution of soft coded vectors as the description of images, and then use maximum likelihood algorithm to estimate the density parameters. Next, the kernel functions between any two images are calculated using a probability product kernel function. Finally, the images are classified by a support vector machine. The experimental results in public image datasets show the proposed algorithm in this paper has achieved better classification performances.

Key words: Bag-of-Features model; statistics; probability product kernels; support vector machine; image classification

0 引言

面对海量的图像信息, 为了合理、高效地对大规模图像进行组织管理, 根据图像的内容自动分类已经成为计算机视觉研究的热点。图像分类的基本过程是首先建立图像内容描述, 然后利用机器学习方法学习图像类别, 最后利用学习得到的模型对未知图像进行分类。在所有图像内容表示方法中, Bag-of-Features (BOF) 模型借用一个视觉词典间接表示

局部特征矢量集合, 将每幅图像表示为视觉单词发生频率的直方图, 缩小了底层视觉特征与高层语义特征之间的“语义鸿沟”, 成为当前流行的图像表示方法^[1]。

BOF 模型的核心思想是利用离线训练好的视觉词典对图像中局部描述矢量进行编码, 编码矢量的统计值作为图像的特征表示。目前 BOF 模型中常用的平均统计值和最大统计值虽然已经表现出优良的图像表征性能, 但是这些统计值只是对编码矢量概率密度分布的粗糙表示^[2], 没有囊括图像中编

收稿日期: 2012-09-12; 修回日期: 2012-12-24

基金项目: 国家自然科学基金重大研究计划项目 (90820306)

第一作者简介: 杨赛 (1981—), 女, 南京理工大学模式识别与智能系统专业博士研究生, 主要研究方向为计算机视觉与机器学习。

E-mail: yangsai6868@163.com

码矢量的全部特征信息。为了解决这一问题,将编码矢量所服从的概率密度分布作为图像内容描述,使用概率乘积核函数度量图像之间的相似性,最后利用基于概率乘积核函数的支持向量机对场景和物体图像进行分类。

1 BOF 模型描述及分析

BOF 模型^[3]的构建过程为:首先提取图像中的局部特征,并使用描述子算法描述局部图像区域,将其中所有训练图像的描述矢量集合记为 $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}$,从集合 $\mathbf{Y}$ 中随机选取 K 个矢量作为初始聚类中心,然后不断进行迭代,算法收敛时的聚类中心为视觉单词,第 j 个视觉单词表示为 $\mathbf{v}_j$, 设 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ 为某幅图像中的局部描述矢量,对其进行硬分配^[4]操作,即

$$u_{ij} = \begin{cases} 1 & j = \arg \min_{j=1, \dots, K} \|\mathbf{x}_i - \mathbf{v}_j\|_2^2 \\ 0 & \text{其他} \end{cases} \quad (1)$$

式中, u_{ij} 为第 i 个编码矢量的第 j 个系数, $\mathbf{x}_i$ 为第 i 个局部矢量。由式(1)可知,硬分配操作将每个局部描述矢量分配给距离最近的视觉单词,会产生很大的量化误差。软分配操作^[5]将局部描述矢量分配给每个视觉单词,能够克服上述缺点,则 u_{ij} 可表示为

$$u_{ij} = \frac{\exp(-\beta \|\mathbf{x}_i - \mathbf{v}_j\|_2^2)}{\sum_{k=1}^K \exp(-\beta \|\mathbf{x}_i - \mathbf{v}_k\|_2^2)} \quad (2)$$

式中, K 为视觉单词数目, β 为光滑参数。经过编码操作后,图像被表示为由 N 个 K 维编码矢量组成的集合,它们的平均统计值计算公式为^[6]

$$\mathbf{B} = \frac{1}{N} \sum_{i=1}^N \mathbf{u}_i \quad (3)$$

式中, $\mathbf{B}$ 为图像的 BOF 特征, $\mathbf{u}_i$ 为每幅训练图像的编码矢量,然后使用支持向量机在此特征空间中对图像分类。BOF 模型中每一个视觉单词为图像组成单元的描述,这些组成单元就是一些中间语义内容,每幅图像都是由这些组成单元组合而成,由此可见 BOF 模型本质上是一种中层语义特征,能够克服图像底层视觉特征与高层语义特征之间的“语义鸿沟”。

BOF 模型中还可以使用编码矢量的最大统计值,最近的研究也表明最大统计值的性能要明显优于平均统计值^[7-8]。为了完全掌握统计方法背后的机制,文献[9-10]对其中的原因进行了解释,认为相

比于平均统计值,最大统计值能够更好地克服背景特征对分类任务的干扰。而实际上图像中的背景特征能够为分类识别提供上下文信息,显然这种解释有所欠缺。从编码矢量在视觉单词向量所张成的空间中服从某一概率分布出发进行解释,编码矢量的统计值可以被看做此概率分布中参数的估计值,最大统计值之所以优于平均统计值,是因为它所对应的似然值更大。由此看来,制约统计方法性能的关键因素在于所使用的统计值并不是分布中参数的最大似然估计值。然而即使使用的统计值恰好与最大似然估计值相同,也只是使用编码矢量的某一特定统计值代替其概率密度分布,会忽略掉编码矢量的高价统计特征,势必会减弱其描述能力,并且各种统计值只能与传统核函数相配合度量图像之间的相似性,例如 RBF 核函数等。在实验部分以软分配操作为例,验证本文给出的关于最大统计值优于平均统计值的原因的正确性,并说明高阶统计值的重要性。为了解决 BOF 模型中使用统计值产生的问题,利用编码矢量所服从的概率密度分布表示图像,使用概率乘积核函数^[11-12]度量图像之间的相似性,最后利用基于概率乘积核函数的支持向量机对图像进行分类。

2 基于概率乘积核函数的图像分类算法

已知两幅图像中的编码矢量 $\mathbf{u}$ 所服从的概率分布分别为 P 与 P' , 根据文献[12]中概率乘积核函数的定义,在概率空间 Ω 中 P 与 P' 之间的核函数为

$$K_\rho(P, P') = \int_{\Omega} P(\mathbf{u})^\rho P'(\mathbf{u})^\rho d\mathbf{u}, \rho > 0 \quad (4)$$

一旦选定核函数,利用支持向量机对图像分类的目标函数为^[13]

$$Q(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K_\rho(P_i, P_j) \quad (5)$$

式中, y_i, y_j 为图像对应的类别标号, α_i 为与每个图像样本对应的 Lagrange 乘子,通过对式(5)中目标函数的最大值求解,得到 l 个支持向量对应的系数 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)$ 不为 0, 支持向量机的分类决策函数为

$$f(x) = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i K(P_i, P) + b^* \right) \quad (6)$$

式中, b^* 为分类阈值,可以用任一个支持向量求得。

则本文算法被应用于图像分类任务中的详细描述如下:

1) 计算每幅训练图像中的局部矢量, 得到的所有局部描述矢量集合记为 $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}$;

2) 对集合 $\mathbf{Y}$ 进行 K 均值聚类, 得到的视觉词典记为 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K\}$;

3) 根据式(1)或式(2)编码每幅图像中的局部描述矢量 $\mathbf{u}_i$;

4) 将任意两幅图像编码矢量的概率分布代入式(4)得到训练图像之间的核函数矩阵 $\mathbf{K}_p$, 基于 $\mathbf{K}_p$ 对式(5)进行求解, 得到支持向量;

5) 对测试图像重复步骤1)~3), 根据式(4)计算测试图像与支持向量之间核函数矩阵, 根据式(6)对测试图像进行分类判别。

由于步骤4)中核函数的计算与编码矢量服从的概率密度分布相关, 因此下面详细推导硬分配和软分配编码矢量对应的概率乘积核函数计算公式的具体表达形式。

2.1 硬分配编码与多项分布

硬分配编码方法将局部描述矢量分配给距离最近的视觉单词, 编码系数只有一个为1, 其余为0, 根据文献[14]中多项分布的定义可知, 编码矢量 $\mathbf{u}$ 满足多项分布, 即

$$p(\mathbf{u} | \boldsymbol{\mu}) = \prod_{k=1}^K \mu_k^{u_k} \quad (7)$$

式中, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$, $\mu_k \geq 0$ 并且 $\sum_{k=1}^K \mu_k = 1$, u_k 为编码矢量 $\mathbf{u}$ 的第 k 维值。将多项分布的表达式代入到式(4), 得到多项分布之间的概率乘积核函数为

$$K(P, P') = \sum_{\mathbf{X}} \prod_{k=1}^K (\mu_k \mu'_k)^{\frac{x_k}{\rho}} = \sum_{k=1}^K (\mu_k \mu'_k)^{\rho} \quad (8)$$

式中, $\mathbf{X} = (x_1, \dots, x_K)$, $\sum x_k = 1$ 。

式(8)是关于多项分布中参数 $\boldsymbol{\mu}$ 的非线性函数, 利用最大似然估计方法计算参数值, 图像中 N 个编码矢量的似然函数为

$$p(\mathbf{D} | \boldsymbol{\mu}) = \sum_{n=1}^N \prod_{k=1}^K \mu_k^{u_{nk}} = \prod_{k=1}^K \mu_k^{\left(\sum_n u_{nk}\right)} \quad (9)$$

式中, $\mathbf{D}$ 为图像数据集。设 $\sum_n u_{nk} = m_k$, 利用 Lagrange 乘子算法, 最大似然函数的求解被转化为

$$\sum_{k=1}^K m_k \ln(\mu_k) + \lambda \sum_{k=1}^K (\mu_k - 1) \quad (10)$$

求解式(10)中关于 μ_k 的导数并设为0, 得到 μ_k 的估计值为

$$\mu_k = -\frac{m_k}{\lambda} \quad (11)$$

将式(11)代入到参数的约束条件中, 得到 $\lambda = -N$, 则多项分布参数 μ_k 的最大似然估计值为

$$\mu_k^{\text{ML}} = \frac{m_k}{N} = \frac{\sum_{n=1}^N u_{nk}}{N} \quad (12)$$

2.2 软分配编码与 Dirichlet 分布

软分配编码方法将局部描述矢量分配给每个视觉单词, 并且编码系数的和为1, 根据文献[14]中 Dirichlet 分布的定义可知: 编码矢量 $\mathbf{u}$ 满足 Dirichlet 分布, 表达式为

$$p(\mathbf{u} | \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^K u_k^{(\alpha_k-1)} \quad (13)$$

式中, $\alpha_k > 0$, $\sum_{k=1}^K u_k = 1$, $B(\boldsymbol{\alpha}) = \frac{\prod_{k=1}^K \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^K \alpha_k\right)}$, 将

Dirichlet 分布的表达式代入到式(4), 得到 Dirichlet 分布之间的概率乘积核函数为

$$K(P, P') = \frac{\int \left[\prod_{k=1}^K u_k^{\alpha_k-1} \right]^{\rho} \left[\prod_{k=1}^K u'_k{}^{\alpha'_k-1} \right]^{\rho} d\mathbf{u}}{B^{\rho}(\boldsymbol{\alpha}) B^{\rho}(\boldsymbol{\alpha}')} = \frac{1}{B^{\rho}(\boldsymbol{\alpha}) B^{\rho}(\boldsymbol{\alpha}')} \int \prod_{k=1}^K u_k^{\rho(\alpha_k + \alpha'_k) - 1} d\mathbf{u} \quad (14)$$

式中, u_k, u'_k 分别为任意两幅图像中编码矢量的第 k 维值, α_k, α'_k 分别为任意两幅图像中的 Dirichlet 分布参数。设 $\beta_k = \rho(\alpha_k + \alpha'_k)$, $k = 1, \dots, K$, 满足 Dirichlet 分布中参数的约束条件, 又因为任何概率密度函数在随机变量整个定义域内的积分为1^[15], 则

$\int \prod_{k=1}^K u_k^{\rho(\alpha_k + \alpha'_k) - 1} = B(\boldsymbol{\beta})$, 将其代入到式(14)中, 得到 Dirichlet 分布之间的概率乘积核函数的表达式为

$$K(P, P') = \frac{B(\boldsymbol{\beta})}{B^{\rho}(\boldsymbol{\alpha}) B^{\rho}(\boldsymbol{\alpha}')} = \frac{\prod_{k=1}^K \Gamma[\rho(\alpha_k + \alpha'_k)] \Gamma\left(\sum_{k=1}^K \alpha_k\right) \Gamma\left(\sum_{k=1}^K \alpha'_k\right)}{\Gamma\left[\sum_{k=1}^K \rho(\alpha_k + \alpha'_k)\right] \prod_{k=1}^K \Gamma(\alpha_k) \prod_{k=1}^K \Gamma(\alpha'_k)} \quad (15)$$

式(15)同样也是关于 Dirichlet 分布中参数的非线性函数, 仍然利用最大似然估计方法计算参数值, 图

像中 N 个编码矢量的对数似然函数为

$$L(D|\alpha) = \sum_{i=1}^N \log p(u_i|\alpha) = N \left[\log \Gamma(\alpha) - \sum_{k=1}^K \log \Gamma(\alpha_k) + \sum_{k=1}^K (\alpha_k - 1) \log \bar{u}_k \right] \quad (16)$$

式中, $\log \bar{u}_k = \frac{1}{N} \sum_i \log u_{ik}$, 由于式(16)没有闭合解析式, 利用定点迭代算法求解上式中的参数^[16], 从 τ 到 $\tau+1$ 步的迭代公式为

$$\alpha_k^{\text{new}} = \psi^{-1} \left(\psi \left(\sum_{k=1}^K \alpha_k^{\text{old}} \right) + \log \bar{u}_k \right) \quad (17)$$

式中, ψ 表示伽玛函数的对数导数函数, 参数的初始值估计公式为

$$\alpha_k^0 = E(u_k) \frac{E[u_l] - E[u_l^2]}{E[u_l^2] - E[u_l]} \quad (18)$$

式中, $l \in \{1, \dots, K\}$, 在计算时可以选用 K 中的任何一个值, 利用式(16)进行迭代计算, 算法收敛时得到 Dirichlet 分布中的参数值。

3 实验

3.1 图像数据集

实验数据集描述如下: 1) 15 类场景数据集^[17]中共有 15 种自然与人工场景, 分别为森林、海滩、山脉、市郊、市区、田野、高速、工厂、街道、建筑物、办公室、商店、厨房、卧室、客厅。每种类别中图像的数目不相同, 图像的平均尺寸为 250×300 , 全部为灰度图像。实验中随机抽取每个类别中的 100 幅图像为训练图像, 剩余图像为测试图像, 同时在训练数据集中随机选取 100 幅图像作为验证集。森林、建筑物、厨房的示例图片如图 1(a) 所示。2) MSRCv2 物体数据集^[18]中有 15 类物体, 每类物体有 30 幅图像,



(a) 15类场景数据集



(b) MSRCv2数据集

图1 数据集的示例图像

Fig. 1 Example images of datasets

图像的平均尺寸为 320×213 , 选择其中的 9 类物体作为实验对象, 分别为飞机、牛、人脸、汽车、自行车、羊、路标、书、椅子。实验中随机抽取每类中 15 幅图像为训练图像, 剩余图像为测试图像, 同时在训练数据集中随机选取 30 幅图像作为验证集。飞机、牛、汽车的示例图片如图 1(b) 所示。

3.2 实验方法及结果分析

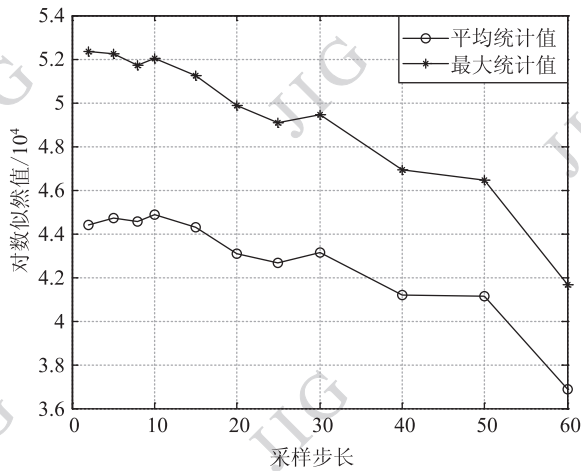
实验环境为: CPU 2.4 GHz, 内存 6 GB, 仿真实验分别从以下几个方面展开。

3.2.1 编码矢量统计值与对数似然函数值

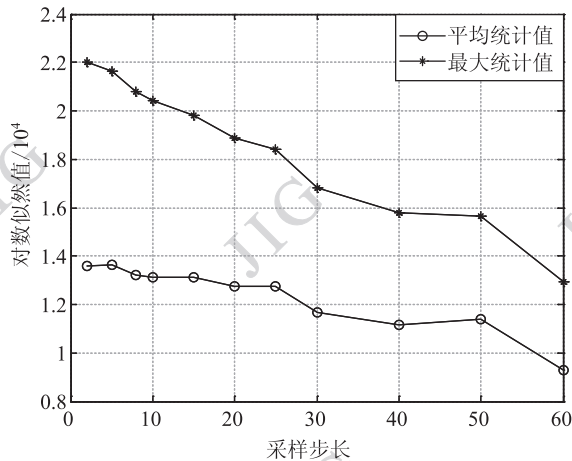
实验对训练数据集中的图像使用划格方法^[19]提取局部区域, 大小为 16×16 , 同时变化采样步长, 分别为 2、6、8、10、20、25、30、40、50、60 个像素值, 使用 SIFT 算子^[20]描述局部图像区域, 使用 K 均值聚类算法构造基于不同采样步长的视觉词典, 视觉单词的数目为 1 500。随机抽取其中某一幅图像作为实验对象, 使用软分配方法对局部描述矢量编码, 计算不同采样步长下图像的平均值与最大值作为 Dirichlet 分布中参数的估计值, 以采样步长为 60 得到的编码矢量作为检验值, 代入到式(16)中, 得到相应的对数似然函数值如图 2 所示。从图 2 可以看到, MSRCv2 物体数据集和 15 类场景数据集取得了一致的实验结果, 在每个采样步长上, 最大统计值对应的对数似然值都要高于平均统计值, 验证了关于最大统计值性能优于平均统计值的原因是正确的。

3.2.2 编码矢量二阶统计值的重要性

实验使用划格方法提取局部区域, 大小为 16×16 , 采样步长为 8 个像素, 使用 SIFT 算子对局部图像区域进行描述, 使用 K 均值聚类构造视觉词典, MSRCv2 数据集上的视觉单词的数目为 1 800, 15 类场景数据集上视觉单词的数目为 1 500。使用软分配方法对局部描述矢量编码, 分别计算编码矢量的平均统计值、最大统计值、二阶统计值作为图像的描述矢量, 使用支持向量机对图像分类, 核函数为 RBF 函数, 利用网格方法在验证集中得到惩罚系数 C 和宽度参数 σ^2 的最优值, 多类分类选用一对一策略, 对应的平均分类正确率以及标准差见表 1。从表 1 可以看到, 在 MSRCv2 物体数据集和 15 类场景数据集上, 使用二阶统计值的图像描述矢量得到的分类正确率要高于使用最大统计值和平均统计值, 说明编码矢量的高阶统计值对于图像的代表性能具有重要作用。



(a) MSRCv2数据集



(b) 15类场景数据集

图2 不同统计值对应的对数似然函数值

Fig. 2 Log-likelihood values of different statistics

表1 不同统计值对应的平均分类正确率

Table 1 Average classification accuracy of different statistics

数据集	平均统计值	最大统计值	二阶统计值
MSRCv2	80.4% (3.7)	88.5% (3.8)	89.9% (3.0)
15类场景	72.3% (0.5)	75.3% (0.3)	76.5% (0.3)

注: 括号内为相应的标准差

3.2.3 与其他图像分类算法的比较

实验中局部特征的提取、局部图像区域的描述以及视觉词典的构造与3.2.2节实验中的方法相同。当使用硬分配方法对局部描述矢量编码时,根据多项分布最大似然估计式(12)得到参数的估计值,根据式(8)计算核函数,然后使用SVM对图像分类,简记为HAPPK;当使用软分配方法对局部描述子编码时,根据Dirichlet分布的最大似然

估计方法,得到参数的估计值,根据式(15)计算核函数,然后使用SVM对图像分类,简记为SAPPK。将本文算法与使用最大统计值的BOF模型进行比较,因此重新实行了硬分配与最大统计值相配合的方法,简记为HAM,软分配编码与最大统计值相配合的方法,简记为SAM;另外,还重新实现了文献[9]算法,简记为LSAM,它是目前性能最优的图像分类算法之一。为了能够在相同条件下进行比较,各种算法均未使用金字塔匹配核函数,因此HAM、SAM、LSAM算法比原文献中报告的平均正确率低6个百分点左右。各种算法对应的平均分类正确率以及标准差见表2。从表2中可看到,在物体与场景分类任务中,对于硬分配编码方法和软分配编码方法,基于概率乘积核函数方法的分类性能要高于相应的使用最大统计值的方法,并且超过了目前性能最优的LSAM算法。各种分类算法的平均运行时间见表3,对于HAPPK算法、SAPPK算法,表中每个数据集对应的第1列为估计编码矢量概率分布参数的平均运行时间,对于HAM算法、SAM算法、LSAM算法,则为计算编码矢量最大统计值的平均运行时间,由结果可知, SAPPK算法的耗时是其他算法的50倍左右,这是因为HAPPK算法估计概率分布参数时,计算 N 个编码矢量的平均值,时间复杂度为 $O(N)$, HAM、SAM、LSAM算法计算编码矢量的最大统计值,时间复杂度也为 $O(N)$,而SAPPK算法需要迭代 t 次估计概率分布参数,时间复杂度为 $O(tN)$;表3中每个数据集对应的第2列为各类算法构建训练图像核矩阵的平均运行时间,由结果可知,各种分类算法在构建核矩阵时所使用的视图相近,这主要是因为HAPPK算法、SAPPK算法中概率分布参数的维数与编码矢量最大统计值的维数相同,都为视觉单词的数目 K ,因此 M 幅训练图像对应的时间复杂度都为 $O(M^2K^2)$ 。

表2 不同分类算法对应的平均分类正确率

Table 2 Average classification accuracy of different algorithms

数据集	HAPPK	SAPPK	HAM	SAM	LSAM
MSRCv2	93.3% (2.2)	93.8% (1.7)	86.8% (2.7)	88.5% (3.8)	88.8% (2.1)
15类场景	75.6% (0.9)	76.8% (0.7)	73.4% (0.4)	75.3% (0.3)	76.0% (0.7)

注: 括号内为相应的标准差

表 3 不同分类算法对应的平均运行时间

Table 3 Average running time of different algorithms

算法	MSRCv2		15 类场景	
	估计编码矢量	构建核矩阵	估计编码矢量	构建核矩阵
HAPPK	0.004 2	5.183 2	0.003 8	9.510 8
SAPPK	0.227 4	6.058 1	0.544 6	10.062 8
HAM	0.004 3	5.167 8	0.004 8	8.423 9
SAM	0.005 9	5.169 3	0.004 0	9.738 6
LSAM	0.004 7	5.153 6	0.004 5	8.914 6

4 结 论

编码矢量的统计值已经成为近年来 BOF 模型研究中的热点。相比于平均统计值,最大统计值虽然大大提升了 BOF 模型的表征性能,但是仍然只是对原本概率密度分布的粗糙表示,并且只能选用传统的核函数度量图像之间的相似性。针对这些问题,将编码矢量所服从的概率密度分布作为图像内容描述,使用概率乘积核函数度量图像之间的相似性,最后利用基于概率乘积核函数的支持向量机对图像进行分类。实验结果表明本文算法的性能超过了目前性能最优的图像分类算法。相比其他图像分类算法,基于硬分配编码的概率乘积核函数方法并没有增加时间复杂度,基于软分配编码的概率乘积核函数方法时间复杂度有所增加,但仍是编码矢量数目的线性函数。由于概率乘积核函数依赖于编码矢量所服从的概率密度分布的具体形式,本文只是给出了硬分配和软分配编码对应的概率乘积核函数的计算公式,因此今后的工作是将概率乘积核函数应用到其他编码方法中并找到适用于所有编码方法泛化的概率乘积核函数。

参考文献 (References)

[1] Yang S, Zhao C X. The image classification algorithm based on latent dirichlet allocation model [J]. Computer Engineering, 2012, 38(14): 181-183. [杨赛, 赵春霞. 基于隐含狄利克雷分配模型的图像分类算法 [J]. 计算机工程, 2012, 38(14): 181-183.]

[2] Liu Y, Perronnin F A. Similarity measure between unordered vector sets with application to image categorization [C] // Proceedings of the 21st International Conference on Computer Vision and

Pattern Recognition. Alaska, USA: IEEE Computer Society, 2008: 1-8.

[3] Csurka G, Dance C R, Fan L X, et al. Visual categorization with bags of keypoints [C] // Proceedings of the 8th European Conference on Computer Vision. Prague, CZE: Springer-Verlag, 2004: 1-22.

[4] Bosch A, Zisserman A, Muñoz X. Scene classification using a hybrid generative/discriminative approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712-728.

[5] Van Gemert J C, Veenman C J, Smeulders A W M, et al. Visual word ambiguity [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(7): 1271-1284.

[6] Wu L, Yu N G. Semantics-preserving bag-of-words models and applications [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 19(7): 1908-1920.

[7] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification [C] // Proceedings of the 22nd International Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Computer Society, 2009: 1794-1801.

[8] Jia Y Q, Huang C, Darrell T. Beyond spatial pyramids: receptive field learning for pooled image features [C] // Proceedings of the 25th International Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE Computer Society, 2012: 3370-3377.

[9] Liu L Q, Wang L, Liu X W. In defense of soft-assignment coding [C] // Proceedings of the 13th International Conference on Computer Vision. Barcelona, Spain: IEEE Computer Society, 2011: 2486-2493.

[10] Boureau Y L, Boch F, LeCun Y, et al. Learning mid-level features for recognition [C] // Proceedings of the 23rd International Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE Computer Society, 2010: 1-8.

[11] Vasconcelos N. On the Efficient evaluation of probabilistic similarity functions for image retrieval [J]. IEEE Transactions on Information theory, 2004, 50(7): 1428-1496.

[12] Jebara T, Kondor R, Howard A. Probability product kernels

- [J]. Journal of Machine Learning Research, 2004, 5: 819-844.
- [13] Zhang X G. Introduction to statistical learning theory and support vector machine[J]. Acta Automatica Sinica, 2000, 26(1): 32-41. [张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-41.]
- [14] Bishop C. Pattern Recognition and Machine Learning[M]. New York: Springer, 2007.
- [15] Rauber T W, Braun T, Berns K. Probabilistic distance measures of the Dirichlet and Beta distributions[J]. Pattern Recognition, 2008, 41: 637-645.
- [16] Huang J. Maximum likelihood estimation of Dirichlet distribution parameters[R]. Pittsburgh: Carnegie Mellon University, 2005.
- [17] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C]//Proceedings of the 19th International Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Computer Society, 2006: 2169-2178.
- [18] Zhang Y, Chen T. Efficient kernels for identifying unbounded-order spatial features[C]//Proceedings of the International Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Computer Society, 2009: 1762-1769.
- [19] Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification[C]//Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer-Verlag, 2006: 490-503.
- [20] Lowe D. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.