基于规范割的空间金字塔图像分类算法

丁 锴 陈伟海 吴星明 刘 中

(北京航空航天大学 自动化科学与电气工程学院, 北京 100191)

摘 要: 对大型图像数据库进行图像分类是很困难的,空间金字塔算法针对这种问题提出,并能得到很好的分类精度,但有几点不足. 针对这些不足,提出基于规范割的空间金字塔算法:使用规范割算法对特征词进行更准确的聚类;对每类训练图像计算子特征库,利用二次聚类生成总特征库,在特征字典中保留更多的稀疏类型图像特征词;用高斯模型量化未知特征生成特征直方图,并对直方图进行尺度重整,提高类间距. 实验证明提出算法比原方法分类精度最多能提高 4.6%.

关 键 词: 数据聚类; 图像分类; 规范割; 支持向量机

中图分类号: TP 391

文献标识码: A **文 章 编 号**: 1001-5965(2013)10-1342-06

SPM based on normalized cut for image classification

Ding Kai Chen Weihai Wu Xingming Liu Zhong

(School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: It is difficult to classify scene images with high accuracy when the dataset is relatively large. Spatial pyramid matching was proposed to deal with this problem, but there are some shortages. As an improvement, the algorithm based on normalized cut was proposed. Normalized cut was utilized instead of *K*-means for clustering. The size of codebook was regulated referring to quantity and size of the images, by calculating sub-codebook for every category and re-clustering the codes. Distance between categories was enlarged by quantifying unknown features with Gaussian model and rescaling the histogram features. Experiments prove that new approach can get higher precision than the original by 4.6% at most.

Key words: data clustering; image classification; normalized cut; support vector machine

随着移动机器人技术的发展,自主性成为机器人性能的一个重要评价标准,机器人视觉无疑是自主性研究最重要的组成部分,图像分类等高层视觉算法成为新的研究热点,并开始使用在智能机器人系统中. 图像分类的研究在早期使用颜色特征和灰度直方图等方法,这些方法受图像中光线、尺度和变形等因素影响比较大. 随着对于尺度和光线不变性研究的进展,主要包括 Harris, SIFT(Scale Invariant Feature Transform)[1] 特征描述子的提出,将这些不变性特征用到图像分类中,产生一些颇受关注的成果,如文献[2-3].

文献[2]提出了词袋 (bag-of-word)算法,feature 在这里指 SIFT 和稠密 SIFT 特征(dense-SIFT). 算法通过统计某些特征在图像中出现的频率,并结合支持向量机等智能算法进行图像分类. 词袋算法对后续研究影响很大,例如文献[4]的 *K* 近邻和文献[5]的空间金字塔(SPM,Spatial Pyramid Matching)及文献[6]的空间金字塔核函数算法等[4-6]及国内的相关工作[7-9],与词袋算法的提出同时,文献[3]提出金字塔匹配核函数(PMK, Pyramid Matching Kernel)算法:首先是在多分辨率情况下,对两幅图像中相同位置的 SIFT 特征点

收稿日期: 2012-11-29; **网络出版时间**: 2013-10-24 14:47

网络出版地址: www.cnki.net/kcms/detail/11.2625.V.20131024.1447.002.html

基金项目: 国家自然科学基金资助项目(61075075,61175108); 北京市科学技术委员会资助项目(D121104002812001)

作者简介: 丁 锴(1983-),男,河南濮阳人,博士生,838383_dingkai@163.com.

进行匹配,并使用加权交叉算法计算相似度[3],然后用最近邻的方法进行分类. 词袋和 PMK 方法对空间信息处理方式迥异,前者统计特征在图像中出现的次数,完全不考虑空间信息,而后者对特征的位置要求严格,只匹配相近的位置上的特征. 融合以上两种方法对空间信息的处理方式,文献[5]提出 SPM 算法,算法先将图片分为若干等面积块,然后在每块上利用词袋算法计算特征,并将特征按图像块顺序排列,这样实现了整体上对位置要求严格,在局部上则放松位置约束,实验结果显示算法的分类准确率有大幅提高.

在对 SPM 的研究中,发现算法有以下问题:①SPM 使用 *K*-means 聚类,此方法容易将数据点错误划分. ②字典(dictionary)生成时,SPM 将所有图像特征一次聚类生成字典,对于纹理稀疏类型的图像,这会使它们对应的特征词在字典中的所占比例很小,增加分类的难度. 针对现有算法存在的问题,研究者们提出一些新的算法,如基于流形的特征词典生成算法[10],组稀疏的特征表示方法[9],都在一定程度上提高了分类效果.

本文提出基于规范割的改进算法,使用规范割算法聚类图像特征,提高聚类精度. 其次使用二次聚类,最大程度的保留纹理稀疏的图像集的相关特征,提高它们与其他类的区分度. 经实验验证,提出算法能够得到更高的分类精度.

1　基于规范割的空间金字塔

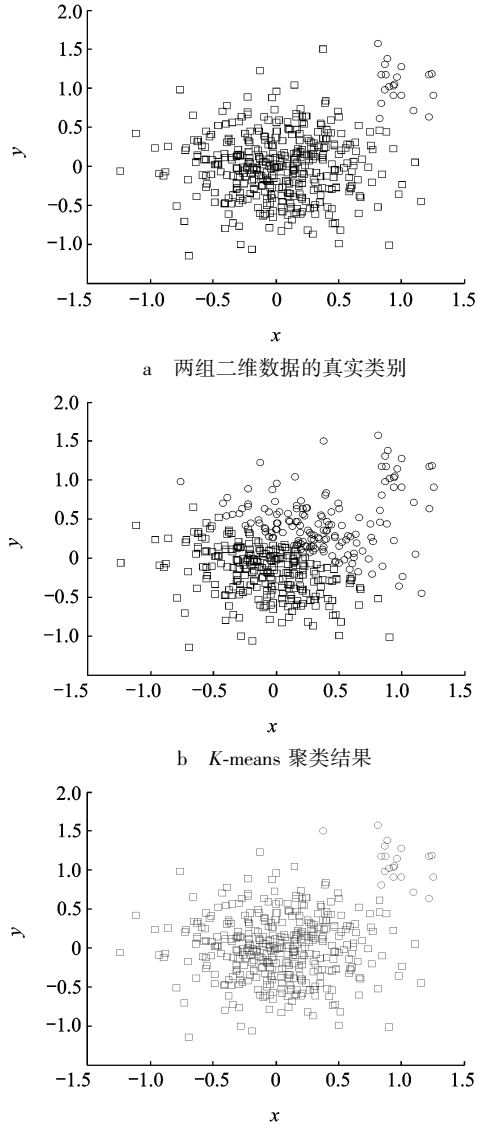
1.1　规范割优化聚类

K-means 算法目的是为了把 n 维空间中的数据集均分到 k 个区域,并使得点心距之和最小,数学表达式为

$$\min_{S}\sum_{\boldsymbol{p}\in\mathbf{R}^n}(\mathrm{dist}(\boldsymbol{p},\boldsymbol{C}))^2 \tag{1}$$

$$\mathrm{dist}(\boldsymbol{p},\boldsymbol{C})=\min_{\boldsymbol{c}\in\boldsymbol{C}}\|\boldsymbol{pc}\| \tag{2}$$

式中,$\boldsymbol{p}$ 表示数据集 $\boldsymbol{S}$ 中任意点;$\boldsymbol{C}$ 表示中心点集. 以二维数据分类为例,随机生成分别以(0,0)和(1,1)为中心的高斯分布点,并用方框、圆圈分别标识,如图 1a 所示;通过 *K*-means 聚类将数据点分为两类,效果如图 1b 所示,可以看出 *K*-means 聚类算法分类出错概率较大;如果对象为样本数量悬殊的多中心数据集,这种分类方法显然是无法使用的. 多种方法可以改善这种情况,例如期望最大化(EM,Expectation-Maximization)算法、规范割等. EM 算法因为计算复杂度原因,无法在图像数据较大的情况下使用. 规范割算法(normalized cut)在近 10 年中发展很快,是一种基于图的方法. 在算法中,由于图的边(edge)值可以表示数据点间的相似关系,所以更容易设定约束条件以优化结果. 规范割对样本数量悬殊的数据有一定鲁棒性,并且效率更高,在此利用此方法实现聚类,结果如图 1c 所示.



a　两组二维数据的真实类别

b　*K*-means 聚类结果

c　规范割聚类结果

图 1　聚类算法示例

规范割[11]将数据的每个点作为图的一个节点,利用点之间的连线权值表示两点的相似度,构造出一个连接图. 如图 2 所示,左右两组数据点 $\boldsymbol{A}$,$\boldsymbol{B}$,中间曲线为割线,图割定义为所有被切割的边值之和:

$$\mathrm{cut}(\boldsymbol{A},\boldsymbol{B})=\sum_{i\in\boldsymbol{A},j\in\boldsymbol{B}}w_{ij} \tag{3}$$

为了防止将孤立点作为一类,规范割最小化函数定义为

$$\mathrm{ncut}(\boldsymbol{A},\boldsymbol{B})=\frac{\mathrm{cut}(\boldsymbol{A},\boldsymbol{B})}{\mathrm{assoc}(\boldsymbol{A},\boldsymbol{V})}+\frac{\mathrm{cut}(\boldsymbol{A},\boldsymbol{B})}{\mathrm{assoc}(\boldsymbol{B},\boldsymbol{V})} \tag{4}$$

$$\left.\begin{aligned}\mathrm{assoc}(\boldsymbol{A},\boldsymbol{A}) &= \sum_{i\in\boldsymbol{A},j\in\boldsymbol{A}} w_{ij}\\ \mathrm{assoc}(\boldsymbol{A},\boldsymbol{V}) &= \mathrm{assoc}(\boldsymbol{A},\boldsymbol{A}) + \mathrm{cut}(\boldsymbol{A},\boldsymbol{B})\end{aligned}\right\} \tag{5}$$

式中，assoc（$\boldsymbol{A},\boldsymbol{A}$）表示集合内的边值之和；assoc（$\boldsymbol{A},\boldsymbol{V}$）表示图上所有点与 $\boldsymbol{A}$ 中节点的边值之和. 当要将数据分为多个类时，最优解方程变为

$$\mathrm{ncut}_k = \frac{\mathrm{cut}(A_1,\boldsymbol{V}-A_1)}{\mathrm{assoc}(A_1,\boldsymbol{V})} + \frac{\mathrm{cut}(A_2,\boldsymbol{V}-A_2)}{\mathrm{assoc}(A_2,\boldsymbol{V})} + \cdots + \frac{\mathrm{cut}(A_k,\boldsymbol{V}-A_k)}{\mathrm{assoc}(A_k,\boldsymbol{V})} \tag{6}$$

求最优解是一个 NP（Non-deterministic Polynomial）完全问题，但是可以使用离散化的方法求得近似解. 定义权值矩阵 $\boldsymbol{W}$，权值也就是两个点之间的相似度，定义为

$$w_{ij} = \exp(-\mathrm{dist}\ (\boldsymbol{p}_i,\boldsymbol{p}_j)/\sigma_k^2) \tag{7}$$

设数据集 $\boldsymbol{S}$ 中有 N 个点，$\boldsymbol{W}$ 为 $N\times N$ 矩阵，σ_k 为第 k 类的类内方差，通过求训练集方差获得. 然后构造对角矩阵 $\boldsymbol{D}$，对角线上每个元素等于对应 $\boldsymbol{W}$ 中每行的加和. 将 Laplace 矩阵定义为 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W}$. 规范 Laplace 矩阵定义为

$$\boldsymbol{L}_{\mathrm{sym}} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2} \tag{8}$$

式中，$\boldsymbol{I}$ 为单位矩阵. 计算规范 Laplace 矩阵的特征值，并按小到大排列，取前 K 个特征值对应的特征向量，得到的 $N\times K$ 矩阵，矩阵每行为一个元素；使每一行的和为 1，转化为矩阵 $\boldsymbol{Q}$. 对 $\boldsymbol{Q}$ 中的 N 个数据进行 K-means 聚类，取 $\boldsymbol{Q}$ 中任一数据 $\boldsymbol{q}_i$，如果 $\boldsymbol{q}_i$ 属于 K 个聚类的某一类，其对应高维映射 $\boldsymbol{p}_i$ 也属于此类. 至此，用规范割求多类聚类问题就解决了. 多类谱聚类方法的推导过程在文献[12]中有详细叙述，其优点是只要求已知相似度矩阵就可求解，并且计算效率高，因为算法将 $N\times N$ 的矩阵压缩到 $N\times K$，然后聚类，通常 $K\ll N$，可以看出这是一种先降维后聚类的方法. 另外谱分析放大了数据间的主要向量上的差别，聚类效果比 K-means 有提升，所以规范割聚类更接近真实值.

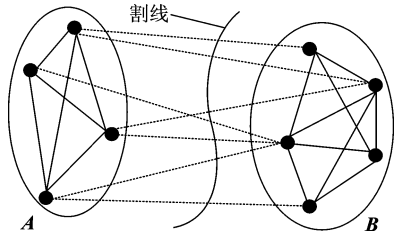


图 2 规范割算法

1.2 二次聚类字典生成算法

SPM 利用聚类算法，将图像集提取的所有 SIFT 特征一次聚类，得到的聚类中心就是特征字典. 此方法的缺点是：由于纹理稀疏类图像空白区域较多，而空白区域不进行聚类，所以稀疏类图像的采样纹理数量比较少，而聚类中心往往向采样纹理集中的特征收敛，纹理丰富类图像占优，稀疏类图像在特征字典中对应元素较少. 本文提出具有自调节能力的二次聚类特征字典生成方法，此方法能在保证计算效率前提下，提高聚类精度. 方法是：先将初始的 K 值按经验设为特征总量的 1%，并对每类图像分别聚类子特征字典（sub-codebook）. 假定一共有 n 类图像，则每类图像的子特征字典大小为 K/n，这种方式使规范割效率更高，因为通常 $K\gg 100$，规范割的效率还不如使用 K-means，并且此聚类方法将纹理稀疏类和稠密类图像分开聚类，解决了聚类中心偏向问题，使特征词在每类中的分布比较均匀. 而后利用均值移动（mean-shift）算法合并子特征字典，生成总字典，即二次聚类. 二次聚类只针对重叠和过离散特征词和它们的相邻特征词，重叠特征词指间距小于某设定阈值的多个特征词，总字典中每两个特征词之间都要有一定距离. 过离散特征词是聚类内部点间距过大的特征词. 利用密度函数评估每个聚类，定义密度函数为聚类内数据点与中心点的距离之和与点个数的比值，找出过离散的聚类. 聚类算法使用均值移动，因为二次聚类处理的数据量不大，且均值移动可以估计真实聚类中心的个数，可将本属于一个聚类，但由于人为设定 K 引起的过度或者不足划分的情况，凝聚或者分离成一定数量的聚类，这样就同时解决了过离散和重叠的特征词的二次聚类问题. 这里，均值移动算法使用由文献[13]提出的数据驱动的自适应确定聚类带宽的方法. 均值移动迭代表达式为

$$\boldsymbol{x}_i' = \frac{\sum_{\boldsymbol{x}_i\in N(\boldsymbol{x})} K(\boldsymbol{x}_i-\boldsymbol{x})\boldsymbol{x}_i}{\sum_{\boldsymbol{x}_i\in N(\boldsymbol{x})} K(\boldsymbol{x}_i-\boldsymbol{x})} \tag{9}$$

式中，$K(\boldsymbol{x}_i-\boldsymbol{x}) = \mathrm{e}^{c\|\boldsymbol{x}_i-\boldsymbol{x}\|^2}$. 算法假设数据符合高斯分布，实验部分使用的数据可以满足此要求. 经过二次聚类生成的特征库中，一些特殊纹理被保存下来，可以提高分类的精度.

1.3 图像表示

图像表示是将图像中提取的 SIFT 特征用特征字典量化并计算直方图的过程，流程如图 3 所示. 在使用特征词量化图像特征时，量化有多种方法，例如用整数集{1,2}量化 1.2，使用 SPM 中的最近距离方法，1.2 就会被量化为 1；使用高斯分布量化方法，由 1.1 节可知，每个特征词对应一个

聚类中心,每个聚类视为一个高斯分布,则聚类中心点为均值,聚类的最大半径为方差,由此可计算每个 SIFT 特征对应特征词的归属概率,选概率最大的特征词为其量化值即可. 使用高斯概率,而不是规范割或者最近邻方法,是因为最近邻方法会引起较大的量化损失,而规范割的方法用作量化的可行性弱. 原因是规范割属于先降维后聚类的方法,降维方法有两种:①将待量化特征和训练集组成矩阵,然后利用特征分解方法降维,此方法每次都要计算训练集组合矩阵的特征向量,计算量大. ②使用训练集的转化矩阵. 转化矩阵是训练集从 $N\times N$ 到 $N\times K$ 降维的映射阵,测试集的样本使用此转化矩阵能直接降维. 此方法看似简单,但问题在于每个图像类都有一个转化矩阵,未知类的转化矩阵存在选择歧义,所以此方法不可行. 使用高斯概率的量化方法,才能保持算法的一致性.

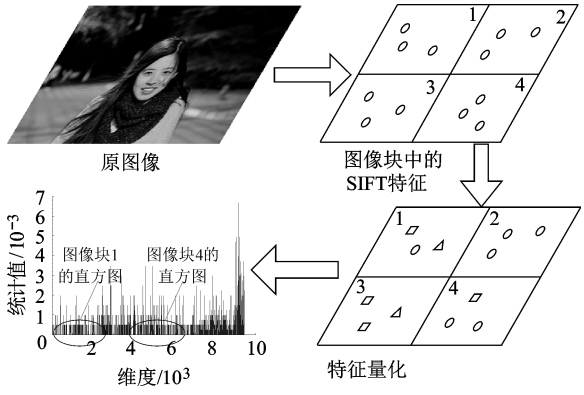


图 3　SPM 图像表示算法流程

经量化后,图像转化为特征直方图矢量,要对特征直方图进行预处理,以放大不同类型之间的距离,并且希望直方图尽量稀疏,从而提高计算效率[14]. 鉴于平滑区域在图像中比例较大,容易在直方图中形成一个很突出的高峰,这种极值能对分类算法产生较大影响,并且平滑区域只是无纹理区域的面积,对类别的区分意义不大,因此应当削弱其带来的影响. 通过阈值判断平滑特征,纹理丰富的区域对应的特征矢量数值波动较大,其方差往往较大,平滑区域对应的特征矢量方差很小. 使用特征字典中最小特征矢量方差作为阈值,小于其值的特征作为平滑特征,在直方图计算时不进行统计.

2　SVM 分类模型

在此,使用线性分类器对数据进行分类. 线性分类通常是对每一类学习一对参数$(\boldsymbol{w}_s,b_s)$,通过$\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{x}+b_s\geqslant>0$ 判定 $\boldsymbol{x}$ 属于 s 类. SVM(Support Vector Machine)是用来线性分类两组数据的,其原理为找到一对平行超平面,使得两组数据分别位于两超平面的不同侧. 满足这种条件的超平面如果存在必然有无限对,而所求超平面对满足间隔最大的限制条件. 分类器表达式为

$$\left.\begin{aligned}&\boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1+\xi_i\\&\min\ \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+P_{\mathrm{c}}\sum_{i=1}^{l}\xi_i\end{aligned}\right\}\tag{10}$$

式中,l 表示数据点数量;$\boldsymbol{x},\boldsymbol{y}$ 表示数据点的坐标;P_{c} 为惩罚参数;ξ 表示位于两平面之间的错误点,当 $\xi>0$ 时,称为柔性分类,这个参数主要用来补偿噪声或者实际数据集中的异常值. 当数据线性不可分时,利用核函数将 $\boldsymbol{x}$ 投影到更高维:

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{\phi}(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}_j)\tag{11}$$

算法使用高斯径向基核函数:

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(-r\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2)\tag{12}$$

式中,r 为核参量,使得分类结果最优的(P_{c},r)参数可以通过交叉验证搜索获得[15]. 这种方法研究得较多,算法比较成熟.

因为 SVM 是二值分类器,所以,针对多类问题可以使用 SVM 的 One-VS-One 分类方式,这种方法对每两类图像训练一个二值分类器,形成分类器组. 每次分类都要通过分析所有二值分类器结果统计数据,利用淘汰法取值作为分类输出. 此方法与 One-VS-All 方法的分类精度相差不多,选任何一种,都不影响分类精度.

3　实　验

在 3 个图像库验证所提出的算法,这 3 个数据库为:15 种场景图像库[5],Caltech-101,Caltech-256,如图 4 所示. 计算所用电脑配置为:CPU 为 Q8300,主频 2.5 GHz,显存 1 G,内存 4 G. 对于 One-VS-One 方法的 SVM,需要训练 $n(n-1)/2$ 个二值分类器,当种类 n 太大时,训练及寻优 SVM 的核参数比较耗时,例如针对 $n=20$ 的分类,程序要持续运行 3 h 以上.

3.1　场景分类

15 种场景库包含 15 类图像,库中每类场景包括 200 ~ 300 张图像,图像大小约为 300 像素×300 像素. 实验设置遵循常用规则,分为两组,分别在每类场景库中随机挑选 20 和 30 张图像组成训练集,另取相同数量图像作为测试集. 特征字典基于训练集生成. 针对第 1 组每类 20 幅图像的情况,子特征字典大小为 40,组成总特征字典并经过优化后大小约为 400. 针对第 2 组每类 30 幅图像的情况,特征字典设置不变. 对两组数据分别进

行 5 次交叉实验,取平均正确率并和原 SPM 算法进行比较. 文献[5]指出,使用每类 100 幅训练图像,特征字典大小为 400 的 SPM,对 15 类场景的最好分类正确率(Precision,下同)为 81.4%,文献[2]对相同大小的训练集正确率为 65.2%. 在 SPM 同样设置情况下,算法得到分类正确率为 81.2%,与 SPM 结果非常接近. 本文在使用 20 及 30 幅训练图像的情况下重新计算分类正确率,原 SPM 算法正确率为 72.7% 及 75.0%. 本文算法对 20 幅训练图像得到的平均正确率为 74.5%,30 幅训练图像正确率为 77.1%. 相对 SPM 整体提高约 2 个百分点,精度略有提升. 分析原因是库中图像特征纹理稠密,平滑区域较少,而且纹理变化繁复,聚类数量真实值较大,不需要再聚类,就能够得到比较好的聚类中心.

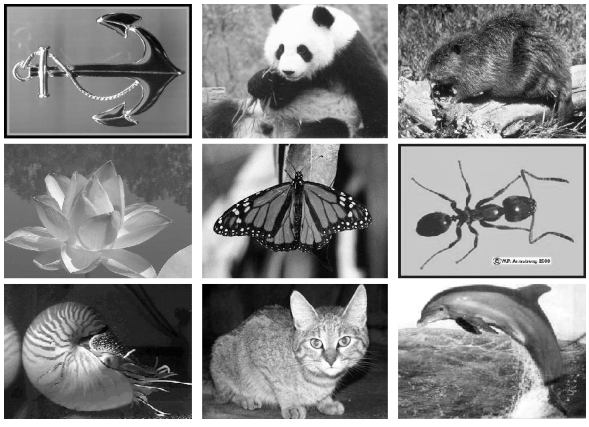
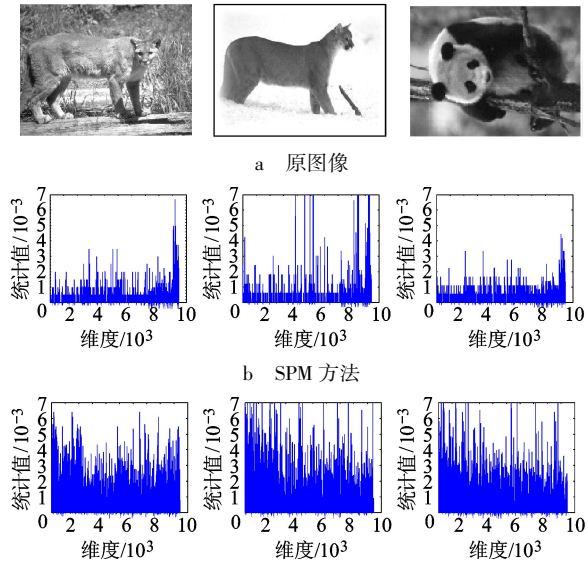


图 4 Caltech-101 图像示例

3.2 物体分类

在这一部分对物体识别进行实验,物体识别跟场景分类相似. 分别对 Caltech-101 及 Caltech-256 数据库进行实验. Caltech-101 共有 101 种类型的图像,包括日常用品,动植物等. 图片大小不均,一般保持长宽比将大图像压缩到 300 像素×300 像素左右,Caltech-256 比 Caltech-101 数据库有更多种类,并且图像的数量和质量都有提高. 数据库中的图像包含物体比较明显,不存在多场景共存的情况,靠人眼很容易识别. 识别率越高,表示算法越好. 同样对 15 类每类选择 20 幅和 30 幅图像的两种训练集进行分类正确率验证,特征字典大小约为 400. 分别使用 SPM 和本文算法进行分类,结果如表 1 所示,本文算法比原 SPM 算法精度高. 生成特征直方图及不同种类的比较见图 5,使用两个直方图特征差值与均值的比值之和作为评价差异度的量化指标,图中两幅美洲狮图像的差异度为:SPM 为 3 491,本文算法为 1 967,第 2 幅与第 3 幅的差异度为:SPM 为 1915,本文算法为 3 564,所以 SPM 分类结果出现错误. 对于 Caltech-256 的图像分类结果如表所示,实验设置与 Caltech-101 相同, Caltech-256 每个图像类型中,可选择的图像更多,图像中物体角度,大小,颜色甚至个体形态都不相同的可能性更大,其类内距必然增大,分类正确率略低.



a 原图像

b SPM 方法

c 本文算法

图 5 特征直方图结果

表 1 针对 15 种图像的分类正确率的比较 %

方法	Caltech-101		Caltech-256	
	20 幅	30 幅	20 幅	30 幅
SPM 算法	68.0	67.8	56.3	57.8
词袋算法	29.3	41.2	30.1	39.7
本文算法	71.0	72.4	59.0	61.8

3.3 精度分析

图像分类的正确率与特征字典大小、类型数量、训练集大小和图像背景等诸多因素有关[2]. 如文献[2]在对 15 种场景库实验时,指出最好的特征字典大小是 172,文献[5]采用此结论并将特征字典大小设定为 200,但在本实验中,400 大小的特征字典正确率更高,文献[16]针对 Caltech 两种数据库分别使用 2 048 及 4 096 大小的特征字典. 由于这些因素的选择不同,这样不同条件下的分类正确率变化很难比较,所以本文将可能影响分类结果的因素都尽量设为一致. 如字典大小设为 400 左右,类型数量 15,训练集图像包括 20 幅和 30 幅两种. 在这种标准下,本文算法有一定提高,特别是对于类似 Caltech 图像库的物体图片,提高较多. 提出的 3 点改进带来不同效果:前两点改进可以看成一点,对算法提高较多. 直方图预处理带来的效果对不同数据库而言是不同的,在纹理比较丰富的场景库中,预处理基本无法改

善分类精度. 本文对其进行的独立实验发现,在场景数据库上引起±0.7% 的精度变化,在纹理比较稀疏的 Caltech 数据库上,改进效果较明显,一般为0.5% ~1.1%. 总体来看,精度提升主要是规范割聚类算法的作用.

4　结　论

空间金字塔算法结合 SVM 能得到很好的场景分类结果,但原算法存在一些问题. 因此,在其基础上提出了改进算法,本文利用规范割算法获得更好的聚类,从而使特征字典更接近最优;对每类图像分别提取特征词,并对重叠和过离散词进行二次聚类,形成优化的字典;通过直方图预处理,剔除无纹理区域特征,以降低其对特征矢量的类别显著性带来的不利影响. 本文用实验验证提出方法比现有类似算法有更好的分类正确率. 但是也可以看出,随着图像类型的增加,维度灾难仍然存在. 由此可知,要实现更高的分类精度,要么添加限制,将图像中的独立目标分离出来;要么增加信息,例如最近提出的分类算法常使用多特征图像表示方法[17]. 所以,未来工作要对多特征图像表示方法及分割算法对图像分类的作用进行研究,从而获得更高的分类精度.

参考文献(References)

[1] Lowe D. Object recognition from local scale-invariant features [C]//Proceedings of International Conference on Computer Vision. Kerkyra: IEEE Computer Society Press, 1999: 1150-1157

[2] Li Feifei, Perona P. A Bayesian hierarchical model for learning natural scene categories [C]//Proceedings IEEE Computer Vision and Pattern Recognition. San Diego: IEEE Computer Society Press, 2005: 524-531

[3] Grauman K, Darrell T. Efficient image matching with distributions of local invariant features [C]//Proceedings IEEE Computer Vision and Pattern Recognition. San Diego: IEEE Computer Society Press, 2005: 627-634

[4] Zhang Hao, Berg A, Maire M, et al. SVM-KNN: discriminative nearest neighbor classification for visual category recognition [C]//Proceedings IEEE Computer Vision and Pattern Recognition. New York: IEEE Computer Society Press, 2006: 2126-2136

[5] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]// Proceedings IEEE Computer Vision and Pattern Recognition. New York: IEEE Computer Society Press, 2006: 2169-2178

[6] Bosch A, Zisserman A. Representing shape with a spatial pyramid kernel [C]// International Conference on Image and Video Retrieval. Amesterdan: Association for Computing Machinery, 2007: 401-408

[7] 张琳波,王春恒,肖柏华,等. 基于 Bag-of-phrases 的图像表示方法[J]. 自动化学报,2012,38(1):46-54
Zhang Linbo, Wang Chunheng, Xiao Baihua, et al. Image representation using bag-of-phrases [J]. Acta Automatica Sinica, 2012, 38(1): 46-54 (in Chinese)

[8] 赵春晖,王莹,Kaneko Masahide. 一种基于词袋模型的图像优化分类方法[J]. 电子与信息学报,2012,34(9):2064-2070
Zhao Chunhui, Wang Ying, Kaneko Masahide. An optimized method for image classification based on bag of words model [J]. Journal of Electronics & Information Technology, 2012, 34(9): 2064-2070 (in Chinese)

[9] 袁莹,邵健,吴飞,等. 结合组稀疏效应和多核学习的图像标注[J]. 软件学报,2012,23(9):2500-2509
Yuan Ying, Shao Jian, Wu Fei, et al. Image annotation by the multiple kernel learning with group sparsity effect [J]. Journal of Software, 2012, 23(9): 2500-2509 (in Chinese)

[10] 刘宝弟,王宇雄,章毓晋. 图像分类中多流形上的词典学习[J]. 清华大学学报:自然科学版,2012,52(4):575-580
Liu Baodi, Wang Yuxiong, Zhang Yujin. Dictionary learning on multiple manifolds for image classification [J]. Journal of Tsinghua University: Science and Technology, 2012, 52(4): 575-580 (in Chinese)

[11] Shi Jianbo, Malik J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905

[12] Ng A, Jordan M, Weiss Y. On spectral clustering: analysis and an algorithm [C]//Advances in Neural Information Processing Systems. Vancouver: MIT Press, 2002: 849-856

[13] Comaniciu D. An algorithm for data-driven bandwidth selection [J]. IEEE Trans Pattern Analysis Machine Intelligent, 2003, 25(2): 281-288

[14] Sarle W. Neural network FAQ [EB/OL]. 1997 [2012-11-20]. ftp://ftp.sas.com/pub/neural/FAQ.html

[15] Hsu Chihwei, Chang Chihchung, Lin Chihjen. A practical guide to support vector classification [EB/OL]. 2002 [2012-11-20]. http://www.cmlab.csie.ntu.edu.tw/~cyy/learning/tutorials/libsvm.pdf

[16] Wang Jinjun, Yang Jianchao, Yu Kai, et al. Learning locality constrained linear coding for image classification [C]//Proc IEEE Computer Vision and Pattern Recognition. San Francisco: IEEE Computer Society Press, 2010: 3360-3367

[17] Yang Jingjing, Li Yuanning, Tian Yonghong, et al. Group sensitive multiple kernel learning for object categorization [C]//Proceedings of International Conference on Computer Vision. Kyoto: IEEE Computer Society Press, 2009: 436-443

(编　辑:刘登敏)