典型相关分析与多伯努利相关模型的图像标注

周 晓1,2, 潘洁珠2

(1. 合肥工业大学 计算机与信息学院,安徽 合肥 230009; 2. 合肥师范学院 计算机科学与技术系,安徽 合肥 230061)

摘 要:文章提出一种基于图像的视觉词袋与文本标注的典型相关分析与分割无关的多伯努利相关模型的自动图像标注算法。在图像标注与分类任务中,矢量量化图像局部描述子得到的视觉词袋特征已显示了其鲁棒性与可区分性,文中对视觉词袋与文本特征作典型相关分析,确保投影变换后新的视觉特征与文本特征的相关性最大化,从而有效地在视觉与文本2种模态中建立联系,契合了自动图像标注的主旨。文中还提出了一种简化的多伯努利相关模型,实验结果证明了典型相关分析比概率潜藏语义分析更适合于图像自动标注,也证明了简化的多伯努利相关模型的有效性。

关键词:自动图像标注;尺度不变特征变换;层次化 k-means;视觉词袋;典型相关分析;多伯努利相关模型中图分类号:TP391.3 文献标志码:A 文章编号:1003-5060(2010)06-0841-06

Automatic image annotation method based on canonical correlation analysis and multiple Bernoulli relevance model

ZHOU Xiao^{1,2}, PAN Jie-zhu²

(1. School of Computer and Information, Hefei University of Technology, Hefei 230009, China; 2. Dept. of Computer Science and Technology, Hefei Normal University, Hefei 230061, China)

Abstract; A novel automatic image annotation method is presented on the basis of canonical correlation analysis between visual bag of words and textual annotation and non-segmentation multiple Bernoulli relevance model. In image annotation or classification tasks, visual bag of words via the vector quantization of image local descriptors often possesses robust and distinct features. In the present method, canonical correlation analysis is made to maximize the correlation between the projected visual features and textual features, thereby to build the relationship between these two modalities. It is the exact intention of automatic image annotation. A simplified multiple Bernoulli relevance model is also proposed, which is another novelty in this paper. Experiment results prove that canonical correlation analysis is better than probability latent semantic analysis in automatic image annotation, and also prove that the simplified multiple Bernoulli relevance model is effective as a new annotation model.

Key words: automatic image annotation; scale invariant feature transform (SIFT); hierarchical kmeans; visual bag of words; canonical correlation analysis(CCA); multiple Bernoulli relevance model (MBRM)

0 引 言

随着数码相机、摄像机及照相手机等移动数 码影像设备的迅速普及,从个人私有媒体库到公 有的万维网,可存取的图像数据正以爆炸的速度 增长。如何有效地为大规模的图像数据建立索引,使得信息真正地对用户是"可见的",随手可及的,是当前图像搜索引擎,如 Google 图片搜索和微软 Live Search, Yahoo! 图像搜索等亟待解决的任务。

收稿日期:2009-05-26;修改日期:2009-09-07

基金项目: 国家重点基础研究发展计划资助项目(2009CB326203);安徽高等学校省级自然科学研究资助项目(KJ2009B238Z) 作者简介: 周 晓(1978-),女,安徽宿松人,合肥工业大学硕士生,合肥师范学院讲师.

图像数据索引的建立,可类似于图书馆里为图书建索引的方式,手工建立文本标注的索引。然而,手工标注大规模的图像集费时费力,且由于个体差异,不同的标注人员可对同一幅图像用不同的关键词描述。所有这些都说明,尽管文本查询方式不失为一种有效的图像检索方法,但用手工标注得到的文本描述不但任务过于繁重,而且非常呆板,缺乏灵活性,并不可取。

针对上述手工标注图像数据的缺陷,近 10 年来产生了一门新的研究领域——自动图像标注,它借助计算机视觉、机器学习与模式识别等多学科技术,从一组已标注好的图像集合中学习图像特征与文本标注 2 种模态间的相关性,从而对新的未知图像,利用这种相关性推理出其最为可能的标注。在过去的几年里,科学家们提出了多种图像标注方法,包括基于机器双语翻译、生成概率模型、图论及多级分类[1-5]等。文献[6]公开的ALIPR 图像自动标注系统原型基本上反映了当前图像标注的发展水平;文献[3]提出了当前性能最高的 EGCap 图像自动标注模型,其创新之处在于提出一种层叠的 LBP 特征,再用典型相关分析对图像特征预处理。

本文提出了一种新的分析图像的视觉词袋特 征,描述与文本特征典型相关的自动图像标注算 法。具体的做法是:首先从图像中提取稠密的彩 色 SIFT 描述子,通过层次化聚类作矢量量化后 得到视觉词汇表,从而把每幅图像用一堆视觉词 汇来表示,又称视觉词袋(visual bag of words)特 征。然后,与传统的采用概率潜藏语义分析 (Probability Latent Semantic Analysis, 简称 PL-SA)来处理视觉词袋特征不同的是,考虑到训练 集的每幅图像所关联的文本特征信息,采用典型 相关分析(Canonical Correlation Analysis, 简称 CCA)把视觉词袋特征与文本标注特征投影到一 个共同的子空间,以最大化两者的内在联系或相 关性,把CCA 投影后的视觉特征作为训练模型 的输入;最后提出了一种分割无关的多伯努利相 关模型(Multiple Bernoulli Relevance Model, 简 称 MBRM)作为自动标注模型。

1 图像的视觉词袋特征描述

图像的视觉特征用于捕捉图像的某种视觉特性,既可以是整幅图像的全局特性,也可以是像素邻域的局部特性。全局特征在对图像的表达上过于简单,现阶段已很少单一采用。局部特征是局

部提取有用的特征描述子,再对其做矢量量化得到图像的综合特征描述,这种方法计算相对复杂,但增加了对图像空间变换的鲁棒性。文献[7]提出的尺度不变特征变换(Scale Invariant Feature Transform,简称 SIFT)是现阶段应用广泛的一种局部描述子。

1.1 SIFT 特征

文献[7]利用尺度空间理论提出了 SIFT 的,构造方法,一般包括 2 个部分:特征提取部分,即 SIFT 检测子;特征表达部分,亦即 SIFT 描述子。

1.1.1 SIFT 检测子

SIFT 检测子决定了 SIFT 特征的几何:几何上,SIFT 特征是个中心为 (x,y),尺度为 σ ,方向为 θ 的方向性圆盘,又称为关键点。这样 SIFT 检测子决定了在图像的哪个位置、哪个方向的多大区域上计算特征的描述子。SIFT 关键点的选取可以是稀疏或稠密 2 种方式^[8]。在图像标注与分类领域广泛采用稠密 SIFT 关键点的选取方式^[9],这里关键点通过对图像栅格化后稠密采样获得,关键点区域的大小可自由设定。虽然这种提取算法看起来过于简单且无严格的数学证明,但是通过该算法检测出的特征在实际的图像分类应用中却被证明也是很有效的^[10]。

1.1.2 SIFT 描述子

SIFT 描述子以梯度的方向直方图的形式描述 SIFT 关键点所在区域的外观,该直方图是关键点区域内梯度方向的统计,如图 1 所示。

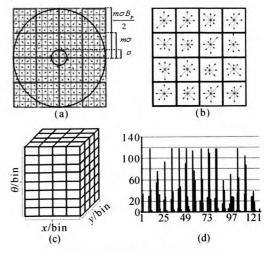


图 1 SIFT 描述子的图例阐述

SIFT 区域块在 x 与 y 方向上量化成 B_p 个 bin, bin 的延拓空间取决于关键点的尺度 σ 大小 (σ 等于图 1a 中小圆圈的半径)与某个放大因子

m(m) 的缺省设置为 3),即 bin 的延拓空间为 mo 像素,这样整个 SIFT 描述子在 x、y 方向上的覆盖空间为 mo B。个像素。此外,描述子把 bin 内像素的梯度方向 θ 量化成 B。级,如图 1b 所示。图 1c 显示 SIFT 描述子是关键点区域内像素的位置(x,y)与梯度方向 θ 的三维直方图 $h(\theta,x$,y),每个像素的梯度都是直方图在(θ ,x,y)上的一个样本,样本以梯度幅值加权,最终累积成直方图 $h(\theta,x,y)$,即为该关键点区域的 SIFT 描述子。实际上,SIFT 描述子还对以关键点为中心的高斯滤波窗加权(以图 1a 中的大圆圈表示),从而分配给距关键点较远像素的梯度幅值更小的权重。作为特征描述子,把上述的三维数组 $h(\theta,x,y)$ 层叠成一维的向量,如图 1d 所示。

1.2 **层次化** k-means 聚类

层次化 k-means 聚类[11] (Hierarchical k-means,简称 Hk-means)首先在整个特征空间中进行 k-means 聚类,将特征空间划分成 k 个子类,此后在每个子类中再分别进行 k-means 聚类,如此循环,不断地将特征空间进行划分,从而可构建一个树状结构:树中每个父节点包含 k 个子节点,树的高度由聚类的层数 L 决定,该过程如图 2a 所示。树的根节点是全部训练图像的集合,树中每个子节点是该层中每个子类的聚类中心。将每个叶子节点看成是一个视觉词汇(Visual Words),亦即特征的聚类中心,便得到了一个词典树,如图 2b 所示。

利用该词典树对图像建立描述的过程,是图像的每个特征沿着词典树的最近路径从根节点向叶子节点走一遍的过程,如图 2c 所示,相应地每幅图像可表示为通过词典树的所有叶子节点的频度向量,即为该图像的视觉词汇的词袋表示。

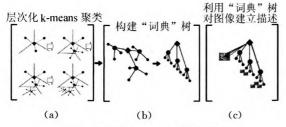


图 2 层次化 k-means 的聚类过程

与 k-means 聚类相比,层次化 k-means 聚类的优势在于聚类速度更快,首先能在较短的时间内建立"词典"树;其次,树状结构大大缩短了对图像建立描述的时间,且可在保证总的词汇数不变

的情况下,通过增加树的深度 L 减小 k 来进一步 地缩短。

2 典型相关分析

典型相关分析由文献[12]提出,文献[13]率 先将 CCA 应用于图像标注方面,此后文献[3]也 延续了这种方法,并取得了显著的实验成果。

在图像标注的问题中,训练集 T 中已标注的图像可表示为图像特征 $\mathbf{x} = (x_1, x_2, \cdots, x_p)^{\mathrm{T}} \in \mathbb{R}^p$ 与文本标注特征 $\mathbf{y} = (y_1, y_2, \cdots, y_q)^{\mathrm{T}} \in \{0, 1\}^q$ 2 组特征向量,这里 p 为图像特征维数,q 为标注词典中的关键词总数(若图像的标注中含有词典的第 i 个关键词,则 $y_i = 1$,反之为 0)。在CCA 投影之前,需对 2 种特征进行归一化处理,即确保特征的每个分量在整个训练集上的期望为 0,方差为 1。

给定归一化后的 x 与 y 向量,CCA 投影过程就是要寻找投影向量 $a_1 = (a_{11}, a_{12}, \cdots, a_{1p})^{\mathrm{T}}$ 与 $b_1 = (b_{11}, b_{12}, \cdots, b_{1q})^{\mathrm{T}}$,得到 x 与 y 投影后的变量 $u_1 = a_1^{\mathrm{T}} x$ 与 $v_1 = b_1^{\mathrm{T}} y$,使得 u_1, v_1 的相关系数 $\rho_1(u_1, v_1) = \frac{a_1^{\mathrm{T}} S_{xy} b_1}{\sqrt{a_1^{\mathrm{T}} S_{xx} a_1 b_1^{\mathrm{T}} S_{yy} b_1}}$ 最大,其中 $S_{xx} \cdot S_{yy} \cdot S_{xy}$ 为 x 与 y 相应的协方差矩阵。为确保解的确定性与唯一性,在以下约束条件下求 $\rho_1(u_1, v_1)$ 的极大值: $a_1^{\mathrm{T}} S_{xx} a_1 = b_1^{\mathrm{T}} S_{yy} b_1 = 1$ 。这样投影向量 a_1 与 b_1 的寻找过程转化为如下极值的求解问题 a_1 ,即

$$\max_{a_1,b_1} (u_1, v_1) = a_1^{\mathrm{T}} S_{xy} b_1,
\text{s. t.} \quad a_1^{\mathrm{T}} S_{xx} a_1 = b_1^{\mathrm{T}} S_{yy} b_1 = 1,
u_1 = a_1^{\mathrm{T}} x,
v_1 = b_1^{\mathrm{T}} y$$
(1)

变量 u_1 、 v_1 称为第 1 对典型变量,类似地可以求得第 2 对投影向量 a_2 与 b_2 及第 2 对典型变量 u_2 和 v_2 ,此时不仅要 ρ_2 (u_2 , v_2)最大,还需确保第 2 对典型变量与第 1 对典型变量的不相关,即 $a_1^T S_{xx} a_2 = b_1^T S_{yy} b_2 = 0$ 。重复上述过程 t 次,即可得到 t 对投影向量,这里 $t \leq \min\{p,q\}$,即仅考虑前 t 个投影向量,类似于主成分分析中的主成分的选取策略。把投影向量作为列向量,最终层叠成 CCA 的 2 个投影矩阵 A 与 B,即

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_t) \in \mathbb{R}^{p \times t},$$

$$\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_t) \in \mathbb{R}^{q \times t}$$
(2)

这样,CCA 投影后的图像特征与文本特征分别为:

$$\mathbf{x}_{\text{CCA}} = \mathbf{A}^{\text{T}} \mathbf{x} \in \mathbb{R}^{t},$$

$$\mathbf{y}_{\text{CCA}} = \mathbf{B}^{\text{T}} \mathbf{y} \in \mathbb{R}^{t}$$
(3)

在图像标注中使用 CCA 的优点主要在于 2 个方面:

- (1) CCA 投影综合考虑了图像与文本标注 两者的相关性,投影后的新的特征向量能更好地 保证图像与文本特征间的一致性,缩短了特征空 间的语义鸿沟。
- (2) 实际上由于只需考虑前 t 对较大的几对 典型变量,CCA 投影后的特征维数 $t \ll \min\{p, q\}$,降维后的特征可有效缩短标注时间。

3 多伯努利相关模型及其简化

3.1 多伯努利相关模型

多伯努利相关模型(MBRM)是图像标注的概率生成语言模型^[2],其基本思想是给定某个测试图像 I,存在潜在的概率分布 P(*|I),亦即 I 的相关模型,在 MBRM 中,相关模型可视为图像 I 中所有可能出现的图像特征 x_i 与文本标注 y 的容器,如图 3 所示。

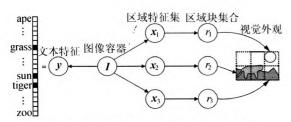


图 3 MBRM 可视为图像外观与标注的生成过程

图 3 中,文本标注 y 是从潜在多伯努利模型 采样得到的二值向量,与此同时,模型还以概率形式决定了图像的特征集 $\{x_1, \dots, x_n\}$,MBRM 又假定图像特征集与图像区域 $\{r_1, \dots, r_n\}$ ——对应,这样最终决定了整幅图像 I 的视觉外观。由此可见 MBRM 依赖于图像分割技术,即把图像 I 分割成 n 个区域 $R_I = \{r_1, \dots, r_n\}$,把每个区域用一个特征来表示,这样图像可表示为一组特征 $X_I = \{x_1, \dots, x_n\}$ 。通过估计其文本标注子集 $Y_I(m_0)$ 为 Y_I 中的标注的数目)与视觉特征 $X_I(n)$ 为特征的个数)的联合分布来标注 I。

3.2 分割无关的简化多伯努利相关模型

由于本文的特征提取过程并不包括图像分割,对每幅图像可用量化后的 SIFT 视觉词袋 CCA 投影后的综合特征来描述,这样本文根据实际情况对 MBRM 算法作如下简化(Simplified MBRM,简称 SMBRM),如图 4 所示。

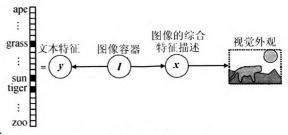


图 4 SMBRM 的生成过程图示

词汇表中任意文本标注 y 与图像特征 x 的联合概率分布为:

$$P(\mathbf{y}, \mathbf{x}_I) = \sum_{J \in T} \{ P_T(J) P_X(\mathbf{x}_I \mid J) P_Y(\mathbf{y} \mid J) \}$$
(4)

其中

$$P_X(\mathbf{x}_I \mid J) = \exp\left\{\frac{-(\mathbf{x}_I - \mathbf{x}_J)^{\mathrm{T}}(\mathbf{x}_I - \mathbf{x}_J)}{2\beta}\right\}$$
(5)

$$P_{Y}(\mathbf{y} \mid J) = (1 - \alpha)\delta_{y,J} + \alpha \frac{n(\mathbf{y}, \mathbf{T})}{|\mathbf{T}|}$$
 (6)

其中,n(y,T)为标注 y 在整个训练集合 T 中的出现次数;|T| 为训练集合 T 中所有图像的标注总数; β 、 α 为 2 个可调参数,分别控制 $P_X(x_I|J)$ 与 $P_Y(y|J)$ 的光滑度 $\delta_{y,J}$,而且有:

$$\delta_{y,J} = \begin{cases} 1, y \in J; \\ 0, y \notin J. \end{cases}$$

4 实验部分

4.1 数据集与性能指标

图像标注的最终性能与训练集和测试集的使用直接相关。尽管存在一定的局限性,仍然采用文献[1]中的 Corel 数据集作为实验数据(简称为Corel5k),这是由于近年来在该数据集上公布的、可供参考的实验成果最为全面,便于从算法性能上进行比较。Corel5k由 5 000 幅等分成 50 类不同主题的彩色扫描图像组合而成,图像由专业摄影师拍摄,同一类图像较为相近,基本上每幅图像由 2~5 个手工标注的词汇描述,词汇表大小为 371。

实验中也使用常用的平均查准率 (average precision)与平均查全率 (average recall)作为图。像标注的性能指标。这里回顾两者的定义:定义 $\sharp \operatorname{Im}_a(y_i)$ 是算法自动给出词汇 y_i 的测试图像的总数, $\sharp \operatorname{Im}_e(y_i)$ 是算法为测试图像正确标注词汇 y_i 的图像总数,而 $\sharp \operatorname{Im}_g(y_i)$ 为在标注真值中含有词汇 y_i 的图像总数。这样查准率与查全率的定义如下:

$$precision(y_i) = \frac{\# \operatorname{Im}_{c}(y_i)}{\# \operatorname{Im}_{a}(y_i)}$$
 (7)

$$\operatorname{recall}(y_i) = \frac{\#\operatorname{Im}_c(y_i)}{\#\operatorname{Im}_g(y_i)}$$
(8)

平均查准率与查全率为:

$$avg_precision(y) = \frac{1}{q} \sum_{i=1}^{q} precision(y_i)$$
 (9)

$$\operatorname{avg_recall}(y) = \frac{1}{q} \sum_{i=1}^{q} \operatorname{recall}(y_i) \quad (10)$$

4.2 实验细节

Corel5k 中的训练集与测试集分别由4500幅和500幅图像组成,实验中又把4500幅的训练集随机划分成5组作交叉认证,每组包括4000幅图像的训练子集与500幅图像的认证集(划分过程中确保认证集中的标注都在训练子集中出现过)。

在特征的选取上,参考文献[9],以稠密的彩色 SIFT 描述子 (Dense Color SIFT, 简称 DC-SIFT) 为特征:以x与y方向上间隔 12 个像素选取关键点,对每个关键点,提取其在 RGB 3 个颜色通道上的 SIFT 描述,并最终串接成一个 SIFT 描述子,其中描述子的参数设置是: B_p =4, B_o =8,bin 的延拓空间分别设置为 $m\sigma$ ={4,6,8,10},即对每个关键点,抽取 4 个不同尺度上的 4×4×8×3=384 维的SIFT 描述子。这样提取训练集中所有图像的 SIFT 描述子,构成 SIFT 描述子集。采用层次化 k-means 算法作特征量化,为每

幅图像建立视觉词汇的词袋表示,得到量化后的 SIFT 词袋特征。实验中,k=3, L=7, 即视觉词汇总数为 $3^7=2$ 187。

对量化后的 SIFT 词袋特征,目前常见的做法是进行概率潜藏语义分析,以学习到一组视觉主题作为新的特征。本文对视觉词袋与文本标注作典型相关分析,得到 CCA 投影变换后新的视觉特征为图像最终的特征描述。这里参数 t=60,即 CCA 投影后特征由 2 187 降至 60 维。实验中为与 PLSA 作公平比较,也取 PLSA 中的参数 K=60,同样可以得到 60 维的主题分布作为新的特征。

最后,对输入的测试图像 I,采用上述相同的方式 提取 CCA 投影后的特征,采用简化的 MBRM 模型计算得到标注结果,即用(4)式计算图像 I 的投影后的特征 x CCA 与词汇表中的每个标注的联合概率,输出前 5 个概率最大的词汇作为 I 的最终标注结果。

实验中需要明确简化的 MBRM 模型中 2 个 参数 β 、 α 的取值,为此采用交叉认证:选择 β 、 α 以最大化 avg_precision(y) + avg_recall(y) 在 5 组 认证 集上 的 均 值,这 样 确 定 的 参 数 值 α = 0.005, β =0.6。

4.3 实验结果比较

图 5 所示对比了本文 CCA-based SMBRM 与文献[3]中 EGCap 的自动标注结果。









dunes, sand, sky, tree frost, fruit, ice buildings, locomotive, lynx, snow 标注真值 smoke, train frost, fruit, ice, frozen, train, railroad, snow, head, bear, polar, water, sky, sand people, **EGCap** locomotive, smoke, bridge frost, ice, fruit, snow, tree CCA-based train, railroad. snow, ice, tree, fox, rocks sand, sky, tree, water, **SMBRM** locomotive, sky, smoke

图 5 图像自动标注的示例对比

表1所列中,在 Corel5k 数据集上,把当今其它优秀算法的性能指标作为基准,比较了本文提出的基于 CCA 与 SMBRM 的图像自动标注系统(简称为 CCA-based SMBRM)的标注性能。用pLSA 替代 CCA,得到第 2 组(简称为 pLSA-

based SMBRM)的标注性能。

首先对比表 1 中的第 2 和第 3 行,从中可以看出,CCA-based SMBRM 比 pLSA-based SMBRM 方案的性能高出 6%左右,由此可以得出结论:在图像标注任务中,CCA 比经典的 pL-

SA性能更优,这是由于CCA兼顾了图像的视觉特征与文本标注2种不同模态的信息描述,这样CCA投影后的新特征能更好地保证图像与文本特征间的一致性,有利于缩短图像标注中最为棘手的语义鸿沟。

表 1 Corel 数据集上不同标注系统的性能比较

方法	平均查准率	平均查全率	备 注
CCA-based SMBRM	0.26	0.27	DCSIFT→Hk-means →CCA→SMBRM
pLSA-based SMBRM	0.19	0.21	DCSIFT→Hk-means →pLSA→SMBRM
EGCap	0.29	0.37	文献[3]
Proba CCA	0.30	0.32	文献[3]
AGAnn	0.24	0.29	文献[4]
SML	0.23	0.29	文献[5]

对比第 2 行与第 4~7 行,可见本文提出的 CCA-based SMBRM 在 Corel5k 数据集上达到了 与 AGAnn 和 SML 相比拟的性能,与 EGCap 算 法相比尚有差距,而在标注模型上,本文的 SMBRM 比 EGCap 更为简单。

5 结束语

本文提出一种新的基于图像的视觉词袋与文本标注的典型相关分析与分割无关的多伯努利相关模型的自动图像标注算法。实验中选取稠密的彩色 SIFT 描述子,通过层次化 k-means 快速聚类后得到图像的视觉词袋特征描述,此后对视觉词袋与文本关键词作典型相关分析,这样得到的CCA 投影后的新特征能更好地保证图像与文本特征间的一致性,有助于缩短语义鸿沟。

此外,本文简化了多伯努利相关模型,提出了一种实现简单的分割无关的标注算法。实验结果证明了 CCA 比 pLSA 更适合于图像自动标注,也证明了本文提出的简化的多伯努利相关模型的有效性[14]。

感谢胡学钢教授与吴共庆老师在论文完成过程中的悉心指导!

[参考文献]

[1] Duygulu P, Barnard K, de Freitas J, et al. Object recognition as machine translation; learning a lexicon for a fixed

- image vocabulary[C]//Johansen P. Proc European Conference in Computer Vision (ECCV'02), Copenhagen, Denmark. Springer, 2002; 97—112.
- [2] Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation[C]//Davis L. Proc Computer Vision and Pattern Recognition (CVPR' 04). Washington, DC: IEEE Computer Society, 2004:1002 —1009.
- [3] Bailloeul T, Zhu C Z, Xu Y. Automatic image tagging as a random walk with priors on the canonical correlation subspace [C]// ACM Multimedia Information Retrieval, Vancouver, Canada, New York; ACM, 2008; 75—82.
- [4] Liu J, Li M, Ma W Y, et al. An adaptive graph model for automatic image annotation [C]//Wang J Z, Boujemaa N. Proc ACM Multimedia Information Retrieval (MIR'06), Santa Barbara, USA: New York: ACM, 2006; 61—70.
- [5] Carneiro G, Chan A B, Moreno P J, et al. Supervised learning of semantic classes for image annotation and retrieval
 [J]. IEEE Trans Pattern Anal Mach Intell, 2007, 29(3): 394—410.
- [6] Li J, Wang J Z. ALIPR: automatic photo tagging and visual image search [EB/OL]. (2006-11-01) [2009-04-30]. http://www.alipr.com.
- [7] Lowe D. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004,60(2):91-110.
- [8] 杨 欣. 基于局部特征的图像分类方法研究[D]. 合肥:中国科学技术大学信息工程学院,2008.
- [9] Bosch A, Zisserman A, Munoz X. Scene classification using a hybrid generative/discriminative approach [J]. IEEE Trans Pattern Anal Mach Intell, 2008, 30(4):712-727.
- [10] Nowak E, Jurie F, Triggs B. Sampling strategies for bag of features image classification [C]//Pinz A. Proc European Conference in Computer Vision (ECCV'06), Graz, Austria. Springer, 2006; 490-503.
- [11] Nister D, Stewenius H. Scalable recognition with a vocabulary tree[C]//Huttenlocher D, Forsyth D. Proc Computer Vision and Pattern Recognition (CVPR'06). New York: IEEE Computer Society, 2006; 2161—2168.
- [12] Hotelling H. Relations between two sets of variants [J]. Biometrika, 1936, 28(3/4): 321—377.
- [13] Hardoon D, Saunders C, Szedmak S, et al. A correlation approach for automatic image annotation [C]//Li X, Zaiane O R, Li Z. Proc Advanced Data Mining and Applications (ADMA'06), Xi'an, China. Springer, 2006:681—692.
- [14] 王雨春. CCA 在数字图像处理中的应用研究[D]. 武汉: 武汉理工大学理学院,2006.

(责任编辑 张秋娟)