基于近义词自适应软分配和卡方模型的图像目标分类方法

赵永威[1],周 苑[2],李弼程[3],柯圣财[3]
(1.武警工程大学电子技术系,陕西西安 710000;2.河南工程学院计算机学院,河南郑州,451191;
3.解放军信息工程大学信息系统工程学院,河南郑州 450002)

摘 要: 传统的视觉词典模型(Bag of Visual Words Model,BoVWM)中广泛存在视觉单词同义性和歧义性问题.且视觉词典中的一些噪声单词—“视觉停用词”,也会降低视觉词典的语义分辨能力.针对这些问题,本文提出了基于近义词自适应软分配和卡方模型的图像目标分类方法.首先,该方法利用概率潜在语义分析模型(Probabilistic Latent Semantic Analysis,PLSA)分析图像中视觉单词的语义共生概率,挖掘图像隐藏的语义主题,进而得到语义主题在某一视觉单词上的概率分布;其次,引入K-L散度度量视觉单词间的语义相关性,获取语义相关的近义词;然后,结合自适应软分配策略实现SIFT特征点与若干语义相关的近义词之间的软映射;最后,利用卡方模型滤除“视觉停用词”,重构视觉词汇分布直方图,并采用SVM分类器完成目标分类.实验结果表明,新方法能够有效克服视觉单词同义性和歧义性问题带来的不利影响,增强视觉词典的语义分辨能力,较好地改善了目标分类性能.

关键词: 视觉词典模型;概率潜在语义分析模型;K-L散度;卡方模型;目标分类
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2016)09-2181-08
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2016.09.024

Image Object Classification Method with Homoionym Based Adaptive Soft-Assignment and Chi-Square Model

ZHAO Yong-wei[1],ZHOU Yuan[2],LI Bi-cheng[3],KE Sheng-cai[3]
(1. *Department of Electronic Technology,Engineering University of CAPF,Xi'an,Shaanxi* 710000,*China*;
2. *Computer College,Henan Institue of Engineering,Zhengzhou,Henan* 451191,*China*;
3. *Institute of Information System Engineering,PLA Information Engineering University,Zhengzhou,Henan* 450002,*China*)

Abstract: The synonymy and ambiguity of visual words always exist in the conventional bag of visual words model based object classification methods. Besides,the noisy visual words,so-called “visual stop-words” will degrade the semantic resolution of visual dictionary. In this article,an image object classification method with homoionym based adaptive soft-assignment and chi-square model is proposed to solve these problems. Firstly,PLSA (Probabilistic Latent Semantic Analysis) is used to analyze the semantic co-occurrence probability of visual words,excavate the latent semantic topics in images,and get the latent topic distributions induced by the words; Secondly,the KL divergence is adopted for measuring semantic distance between visual words,which can get semantically related homoionym; then,adaptive soft-assignment is proposed to realize the soft mapping between SIFT features and some homoionym; finally,the Chi-square model is introduced to eliminate the “visual stop-words” and reconstruct the visual vocabulary histograms,and moreover,SVM (Support Vector Machine) is used to accomplish object classification. Experimental results indicated that the adverse effects produced by the synonymy and ambiguity of visual words can be overcome effectively,the distinguishability of visual semantic resolution is improved,and the image classification performance is substantially boosted compared with the traditional methods.

Key words: bag of visual words model; probabilistic latent semantic analysis; K-L divergence; Chi-square model; object classification

---

收稿日期:2014-12-03;修回日期:2015-03-09;责任编辑:梅志强
基金项目:国家自然科学基金(No.60872142,No.61379152);全军军事学研究生课题资助项目(No.YJS1062)

1 引言

随着计算机技术、通信技术的飞速发展及广泛应用,形成了海量图像信息环境.如何让计算机对其进行快速有效的分类处理,已成当前计算机视觉领域亟待解决的问题.视觉词典模型(Bag of Visual Words Model,BoVWM)[1~5]已成为目前图像目标分类领域[6]的主流处理方法.其基本思想是利用 K-Means 等聚类算法[7,8]对训练图像库中提取的局部特征(通常选取 SIFT 特征[9])集合进行聚类生成视觉码本,也即视觉词典,然后,将每幅图像的 SIFT 特征与视觉词典进行映射匹配得到表征图像内容的视觉词汇直方图,最后,结合机器学习方法训练识别测试图像的类别.然而,由 K-Means 及其改进聚类算法生成的视觉词典存在视觉单词同义性和歧义性问题[10].

为了克服视觉单词同义性和歧义性问题带来的不利影响,研究人员进行了诸多尝试.Philbin 等[11]提出了一种基于软分配的视觉词典模型方法(Soft Assignment,SA)来构建视觉词汇分布直方图,将一个 SIFT 特征分配至与之距离最近的几个视觉单词上,并根据距离大小赋以相应的权重.Gemert 等[10]提出了视觉单词不确定性(Visual Word Uncertainty)模型,通过核函数完成图像局部特征点与视觉单词之间的软映射,有效地减小了特征点与视觉单词映射匹配时的量化误差.Koniusz 等[12]则进一步验证了软分配方法对克服视觉单词同义性和歧义性,减小量化误差的有效性.Li 等[13]在构建直方图时引入了一种上下文信息的策略提高了特征点与视觉单词间的匹配精度,在一定程度上降低了单词同义性和歧义性导致的量化误差.Weinshall 等[14]则将软分配策略与潜在狄里克雷分布模型相结合(Latent Dirichlet Allocation,LDA),提出了种一软分配的 LDA 模型;Danilo 等[15]考虑到视觉单词歧义性的影响,提出了一种模糊聚类的算法完成视觉单词的软分配,并取得了不错的效果.上述方法较于传统的硬分配的视觉词典模型方法[8](Hard-Assignment,HA)都能在一定程度上克服视觉单词的同义性和歧义性问题,减小特征与单词映射时的量化误差,增强视觉词汇直方图特征的语义表达能力.但是,它们都以特征空间距离大小来衡量单词之间的语义距离大小,而由于度量空间的不一致性,使得特征空间距离相近的视觉单词在语义空间并不一定相近.此外,这些方法[10~15]在软分配时对每个局部特征都分配相同数量的视觉单词,难免会使一些不具有歧义性的局部特征也都强制性的映射到了多个视觉单词上,引入新的噪声和冗余信息.

此外,由于图像背景噪声的存在和聚类算法的局限性[16,17],使得生成的某些视觉单词类似于文本信息中的“的”、“和”、“是”等“停用词”,从而降低视觉词典的语义分辨能力,这里称其为“视觉停用词”.Sivic 等[1]考虑到单词的信息量大小与其出现的频率有一定的关系,从而提出了一种基于词频的“视觉停用词”过滤方法.Yuan 等[18]试图以统计视觉单词组合也即“停用词组”出现的概率来滤除一些无用信息,但是却忽略的视觉词组内部各单词的顺序.Chen 等[19]则提出了一种强分辨力的视觉词组(Discriminative Visual Phrases,DVP)筛选方法,在滤除噪声的同时有效克服了传统视觉词组构建方法[20]导致的特征信息丢失问题.然而,这几种方法都忽略了视觉单词和图像类别和语义概念间的相互关系,容易错误地将一些出现次数较少而分辨力较强的视觉单词当作“视觉停用词”.

综上所述,为了更加准确地衡量视觉单词间的语义相关性,且针对不同类别的局部特征自适应地选择软分配数目,同时,有效滤除“视觉停用词”.本文提出了一种基于近义词自适应软分配和卡方模型的图像目标分类方法,解决视觉单词同义性和歧义性问题及其带来的不利影响,增强视觉词典的语义分辨能力,进而提高目标分类准确率.

2 基于近义词自适应软分配和卡方模型的图像目标分类

对于训练图像集 $C=\{C_1,C_2,\cdots,C_k\}$,这里,采用文献[9]的方法提取 SIFT 特征,并采用近似 K-Means 算法[11](Approximate K-Means,AKM)对特征点聚类生成视觉词典.基于近义词软分配和卡方模型的图像目标分类方法具体流程如图 1 所示.首先,通过 PLSA 分析图像中视觉单词的语义共生概率,挖掘图像潜在的语义主题,进而得到语义主题在某视觉单词上的概率分布;然后,引入 K-L 散度度量视觉单词间的语义距离,得到语义相近的近义词,并根据 SIFT 的模糊性自适应地选择软分配视觉单词数目,实现 SIFT 特征与若干语义相近单词之间的软映射;最后,采用卡方模型分析视觉单词与各图像类别之间的相关性,滤除若干相关性小的“视觉停用词”,重构视觉词汇直方图,并由 SVM 分类器完成图像目标分类.

2.1 视觉单词语义概念表达与度量

传统的通过计算单词间欧氏距离来衡量视觉单词间语义距离的方法[10,14],并不能准确地诠释单词间的语义相关性.文献[13]通过获取图像类别在视觉单词上的条件概率分布来代表单词所表达的语义概念,取得了较好的分类效果,但是该方法的前提是来自不同类别的图像中不能包含相同的语义概念.而通过 PLSA 模型能够获取语义主题在某一视觉单词上的条件概率分布,能更为准确地表达单词蕴含的语义概念.下面介

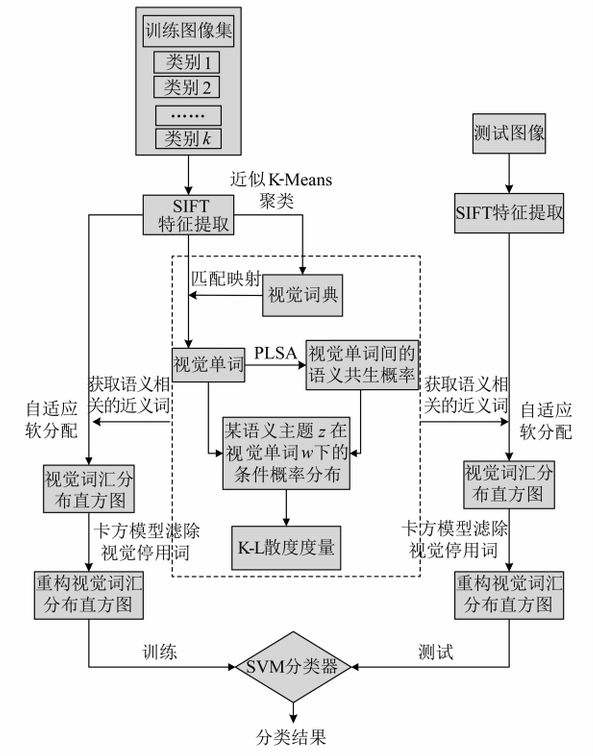

图1　基于近义词自适应软分配和卡方模型的图像目标分类方法流程

绍基于 PLSA 模型的视觉单词语义概念表达.

2.1.1　基于 PLSA 模型的视觉单词语义概念表达

PLSA 模型是由 Hoffman 等[21]针对潜在语义分析提出的一种主题生成模型. 其关键思想是通过分析已知文档集中单词的共生概率,学习得到模型参数,进而预测未知文档隐藏的主题 $z_k(k=1,2,\cdots,K)$,后被广泛应用于计算机视觉领域的图像潜在主题分析. 对于训练图像集 $C=\{C_1,C_2,\cdots,C_k\}$ 和由 AKM 聚类生成的视觉词典 $W=\{w_1,w_2,\cdots,w_n\}$,一幅图像 I 的潜在语义主题 z 分析过程可描述如下:

(1)选择一幅图像,得其观测概率 $P(I)$,$P(I)$ 表示在训练图像库中观测到图像 I 的概率;

(2)选择一个潜在主题 z,得 $P(z|I)$,$P(z|I)$ 表示主题 z 在图像 I 下的条件概率分布;

(3)在已知主题的条件下,得到单词 w 的条件概率 $P(w|z)$.

重复上述过程就能得到图像和视觉单词的共现频率矩阵 $\boldsymbol{N}=[n(w_i,I_j)]$,其中 $n(w_i,I_j)$ 表示图像 I_j 中单词 w_i 出现的次数. 那么,(w,I) 的联合分布可计算如下:

$$\begin{aligned}P(w,I)&=P(I)P(w|I)\\&=P(I)\sum_{z\in Z}P(w|z)P(z|I)\\&=\sum_{z\in Z}P(w|z)P(z)P(I|z)\end{aligned}\tag{1}$$

其中,Z 表示潜在语义空间中所有的主题集合. 而根据最大似然准则,变量 $P(z)$,$P(w|z)$,$P(I|z)$ 可以通过 EM 算法迭代式(2)的最大化对数似然函数得到.

$$\begin{aligned}&L=\lg P(C,W)=\sum_{I\in C}\sum_{w\in W}n(w,I)\lg P(w,I)\\&\text{s.t.}\ \sum_{z\in Z}P(z)=1,\sum_{w\in W}P(w|z)=1,\sum_{I\in C}P(I|z)=1\end{aligned}\tag{2}$$

然后,再利用贝叶斯估计就能得到单词 w 的出现概率和主题 z 在 w 下的条件概率分布,如式(3)和式(4)所示:

$$P(w)=\sum_{z\in Z}P(w|z)P(z)\tag{3}$$

$$P(z|w)=\frac{P(z,w)}{P(w)}=\frac{P(w|z)P(z)}{\sum_{z\in Z}P(w|z)P(z)}\tag{4}$$

然而,在当前 PLSA 模型中主题数目的多少大多是由人工根据经验设定的一个固定值[22],并在此基础上训练主题模型,得到固定主题集下的图像语义表示. 这种人工设定主题数的方法忽略了各图像类别之间内容繁简不一的情况. 为此,可以采用文献[23]中的基于密度的最优 PLSA 模型主题数选择方法,该方法在为各图像类别语义内容构建主题模型时,能够依据图像内容复杂度较好地自动设置语义主题数.

2.1.2　基于 K-L 散度的语义距离度量

K-L 散度[13]可以很好地用来衡量两个概率分布之间的差别,因此得到主题 z 在 w 下的条件概率分布后就能引入 K-L 散度度量不同视觉单词间的语义距离. 而同一幅图像可能会包含多个潜在语义主题,且不同的语义主题对表达图像语义内容的贡献是不一样的,因此,需要对不同的语义主题分配不同权重. 研究表明,训练集 C 在某一主题下 z 的条件熵 H 能够衡量某一语义主题 z 的分辨力. H 值计算如下:

$$\begin{aligned}H(C|z\in Z)&=P(z)\sum_{I\in C}P(I|z)\lg\left(\frac{1}{P(I|z)}\right)\\&=-P(z)\sum_{I\in C}P(I|z)\lg(P(I|z))\\&=-\sum_{I\in C}P(I,z)\lg(P(I|z))\end{aligned}\tag{5}$$

从式(5)中不难看出,条件熵 H 的值越大,数据集 C 的表达内容的不确定性就越大,也即是该主题 z 的分辨力较弱,为此,采用式(6)对条件熵值 $H(C|z)$ 进行高斯归一化得权值 $\omega(z)$ 表征主题 z 的贡献大小.

$$\omega(z)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(1/2)H^2(C|z)}\tag{6}$$

然后,就可以利用 K-L 散度度量两视觉单词 w_i,w_j 之间的语义距离,如式(7)所示:

$$d(w_i,w_j)=\mathrm{KL}(P(z|w_i)\parallel P(z|w_j))$$

$$= \sum_{z \in Z} \omega(z) P(z|w_i) \lg \frac{P(z|w_i)}{P(z|w_j)} \tag{7}$$

不难看出,式(7)在计算视觉单词 w_i,w_j 间语义距离时,K-L 散度同时考虑了主题 z 的权值,但 K-L 散度是一个非对称的距离度量,也即是并不能保证 $d(w_i, w_j) = d(w_j, w_i)$. 为此,对其式(7)改进如式(8)所示,使其为一个标准的对称式距离度量.

$$\begin{aligned} d(w_i, w_j) = & \\ & \frac{P(w_i)}{P(w_i) + P(w_j)} \cdot \mathrm{KL}(P(z|w_i) \parallel P(z|w_j)) \\ & + \frac{P(w_j)}{P(w_i) + P(w_j)} \cdot \mathrm{KL}(P(z|w_j) \parallel P(z|w_i)) \end{aligned} \tag{8}$$

由式(3)~式(7)就能计算两个视觉单词间的语义距离,获取语义相关的近义词,进而结合软分配策略构建视觉词汇分布直方图,更好地克服视觉单词同义性和歧义性带来的不利影响.

2.2 自适应软分配构建视觉词汇直方图

由 PLSA 模型及 K-L 散度度量得到语义相关的近义词之后,若要实现自适应软分配构建视觉词汇分布直方图,首先需要对 SIFT 特征的模糊性进行分析,其模糊性示意图如图 2 所示. 其中,圆点代表 SIFT 特征,椭圆代表视觉单词,菱形和正方形代表两种不同性质的 SIFT 特征. 对于菱形特征而言,与视觉单词 w_1 的距离最近,且与其他视觉单词距离较远,则可假定其代表的语义内容可由视觉单词 w_1 来表达. 也即是该特征点不具有模糊性或者模糊性很小,定义这类可靠特征为第一类特征;对于正方形特征而言,其距离视觉单词 w_2 和 w_3 之间(或者与更多单词之间)的距离很近,则可假设其代表的语义内容需由 w_2 和 w_3 或更多视觉单词共同来表达. 也即是该特征点具有较大的模糊性,定义这类模糊特征为第二类特征.

自适应软分配即是对图像中每一个 SIFT 特征到近义词之间的距离进行分析和归类,然后对不同类别的 SIFT 特征采用不同的的分配策略. 假设已经建立好的视觉词典为 $W = \{w_1, w_2, \cdots, w_n\}$,其中,$n$ 为视觉词典规模大小. 那么,不同类别的 SIFT 特征,就能自适应地将其映射到一定数目的视觉单词上. 过程可描述如下:

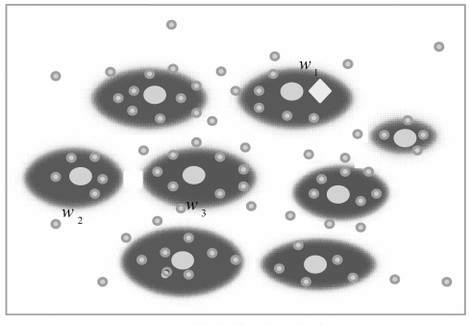

图2 SIFT特征模糊性示意图

(1)对于任一幅图像 $I = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_i, \cdots, \boldsymbol{r}_T\}$,$T$ 表示图像 I 中 SIFT 特征点数目,找到视觉词典 W 中与 SIFT 特征 $\boldsymbol{r}_i$,$1 \leqslant i \leqslant T$ 距离最近的视觉单词 w_k^1;

(2)根据 2.1 节中式(8)计算得到视觉词典中与视觉单词 w_k^1 语义距离最近的 m 个视觉单词 w_k^j,$(1 \leqslant j \leqslant m)$(注意这 m 个单词中包含 w_k^1 自身),然后,分别计算 SIFT 特征与 m 个单词之间的距离,并按从小到大的顺序进行排序,记为 $d = \{d_1, d_2, \cdots, d_j, \cdots, d_m\}$,其中,$d_j$ 表示单词与特征点相距第 j 近的距离;

(3)然后,根据准则 $N_{\mathrm{adp}} = \arg\max_i \{d_i \leqslant \alpha \cdot d_1\}$,$(i = 1, 2, \cdots, m)$ 自适应地判定特征点 $\boldsymbol{r}_i$ 映射时需要分配的视觉单词个数 N_{adp}. 并根据"距离越近关系越密切"的原则,按照 $\mathrm{e}^{-d_l^2/(2\sigma^2)}$ $(l = 1, 2, \cdots, N_{\mathrm{adp}} \leqslant m)$,重新分配该单词的权重. 其中,$\alpha$ 为"自适应软分配因子",通常是一个大于等于 1 的数值,用来控制分配数目.

2.3 "视觉停用词"滤除

传统的"视觉停用词"滤除方法主要是依据词频高低,这种方法容易出现误判现象. 而卡方模型[24]是一种常用的测量两个随机变量独立性的方法,利用卡方模型能够统计视觉单词与各图像类别之间的相关性,卡方值越小表示该视觉单词与各图像类别的相关性越小,区分性也就弱,反之亦然. 因此,可以在统计单词词频的基础上,结合卡方模型更好地滤除"视觉停用词". 这里,假设视觉单词 w 的出现频次独立于图像类别 C_j,$C_j \in C$,$1 \leqslant j \leqslant k$,训练图像集 $C = \{C_1, C_2, \cdots, C_k\}$,而视觉单词 w 与图像集 C 中图像类别的相互关系可以由表 1 来描述.

表 1 视觉单词 w 与各目标类别的统计关系

	C_1	C_2	…	C_k	Total
包含 w 的图像数目	n_{11}	n_{12}	…	n_{1k}	n_{1+}
不含 w 的图像数目	n_{21}	n_{22}	…	n_{2k}	n_{2+}
Total	n_{+1}	n_{+2}	…	n_{+k}	N

表中,n_{1j} 表示图像类别 C_j 包含单词 w 的图像数目,n_{2j} 表示图像类别 C_j 不包含单词 w 的图像数目,n_{+j} 则表示图像类别 C_j 中的图像总数,并用 n_{i+},$i = 1, 2$ 分别表示图像集 C 中包含单词 w 的图像总数和不包含 w 的图像总数. 那么,表 1 中视觉单词 w 与各图像类别的卡方值可计算如下:

$$x^2 = \delta = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(Nn_{ij} - n_{i+}n_{+j})}{Nn_{i+}n_{+j}} \tag{9}$$

卡方值的大小则表征了视觉单词 w 与各图像类别之间的统计相关性大小,同时为了考虑单词词频的影响,这里为每个视觉单词的卡方值赋予相应的权重如下:

$$\tilde{x}^2 = \frac{x^2}{\mathrm{tf}(w)} \tag{10}$$

其中,$\mathrm{tf}(w)$ 表示单词 w 的词频. 不难看出,式(10)同时兼

顾了视觉单词 w 的词频及其与各图像类别之间的统计相关性,因而能更准确地判别单词 w 是否为"视觉停用词".通常的做法是按照式(10)对单词的卡方值进行排序,然后去除一定数量 S 的"视觉停用词"即可,而在重构视觉词汇分布直方图时,对应单词的维度将被滤除.

3　实验结果与性能分析

3.1　实验设置与性能评价

实验数据采用目标分类常用的 Caltech-256 图像集和和 Proval Voc 2007 数据集[25]对本文方法性能进行评估. 随机选取 Caltech-256 图像集中的 15 个目标类别进行实验以验证文中各方法的有效性. 并从每个类别中随机选取 50 幅,共 750 幅图像构成训练图像集,其余作测试集,视觉词典规模为 1000. 图 3 给出了每个目标图像示例. 这里分类采用的 SVM 分类器,具体为 LIBSVM[26]工具包,其核函数采用径向基型内积函数. 而为了获取可靠的实验结果,所有结果都是进行 10 次独立的目标分类实验平均得来. 实验硬件配置为一台 Core 3.1G ×4 CPU,内存为 4G 的台式机. 目标分类性能评价指标为召回率、准确率,以召回率为基础的混淆矩阵(Confusion Matrix)以及平均准确率(Average Precision,AP),相关定义如下:

$$\text{召回率} = \frac{\text{被正确分类的图像数}}{\text{该类图像总数}} \times 100\% \quad (11)$$

$$\text{准确率} = \frac{\text{被正确分类的图像数}}{\text{被分类的图像总数}} \times 100\% \quad (12)$$

$$\text{AP} = \frac{\text{各图像类别分类准确率之和}}{\text{图像类别总数}} \times 100\% \quad (13)$$

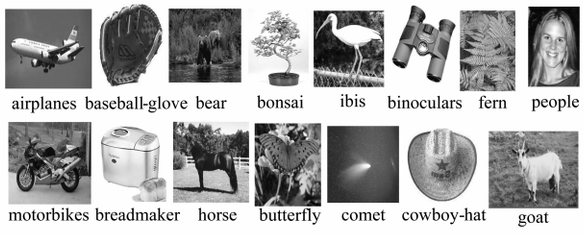

图3　数据集各类别图像示例

3.2　实验结果与分析

首先,为了验证文中基于 PLSA 模型的近义词软分配方法(PLSA + Soft Assignment,PLSA + SA)对克服视觉单词同义性和歧义性问题的有效性,将其与传统的软分配方法[12](SA)和硬分配方法[8](HA)相比较,得到三种方法随单词软分配数目变化的分类平均准确率 AP 值,如图 4 所示. 从图 4 中可以看出,SA 方法及本文的 PLSA + SA 方法的分类准确率均高于 HA 方法. HA 方法的 AP 值始终保持在 66.3%. SA 方法及 PLSA + SA 方法的 AP 值则先是随单词软分配数目的增大而增大,当软分配数目超过一定数目时,准确率反而呈一定的下降趋势,且软分配数目大于 7 时,SA 的分类效果反而差于 HA 方法. 而本文中的 PLSA + SA 方法由于能够从语义概念表达上分析单词间的相似性,进而将相应特征点分配至若干与之语义相近的视觉单词上,可以更准确地表达图像内容,其分类准确率也优于传统 SA 方法.

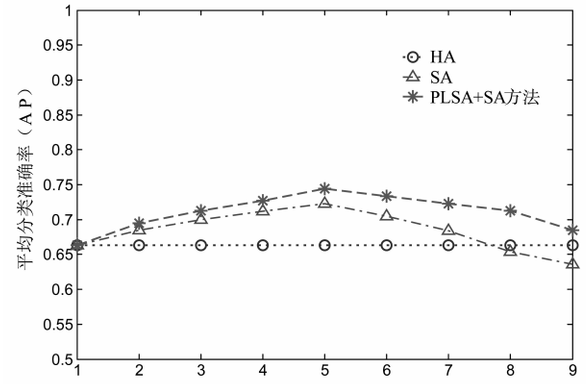

图4　不同分配方法的AP值对比

需要注意的是,在图 4 中的实验中为每个 SIFT 特征点都分配了相同的单词数目,并没有考虑不同 SIFT 特征之间的差异性,难免会使一些不具有歧义性的局部特征也都强制性的映射到了多个视觉单词上,引入新的噪声和冗余信息. 由 2.2 节内容可知,通过分析 SIFT 特征的模糊性类别进而实现自适应软分配的方法能够在一定程度上克服该问题. 因此,为了验证这种自适应软分配的效果,并分析其随自适应软分配因子 α 的变化情况. 在利用 PLSA 模型得到近义词之后分别采用传统的软分配方法(即 PLSA + SA)和自适应软分配方法(PLSA + Adaptive Soft-Assignment,PLSA + ASA)进行分类实验,令 2.2 节中自适应软分配方法中的 m = 20,且 PLSA + SA 方法的 AP 值选择的是单词软分配数目为 5 时的 74.4%. 得目标分类的 AP 值如图 5 所示. 从图 5 中可以看出,随着参数 α 的增大,具有不同模糊类别的 SIFT 特征能够更准确地分配到若干近义词上,PLSA + ASA 方法的分类平均准确率也随之提高,当 α =2 时,取得最高 AP 值 77.86%,优于 PLSA + SA 方法. 然而,当 α 值增大到一定程度时,其分类 AP 值会呈一定的下降趋势,因为过大的 α 值同样会引起传统软分配方法导致的过分配问题. 需要注意的是 α 的取值与训练数据密切相关.

而为了验证文中卡方模型滤除"视觉停用词"的效果,实验将基于近义词自适应软分配与卡方模型相结合(PLSA + ASA + CSM)验证过滤不同数目"视觉停用词"对分类结果的影响,并与未进行视觉停用词滤除时的目标分类结果进行对比,得其分类准确率如图 6 所

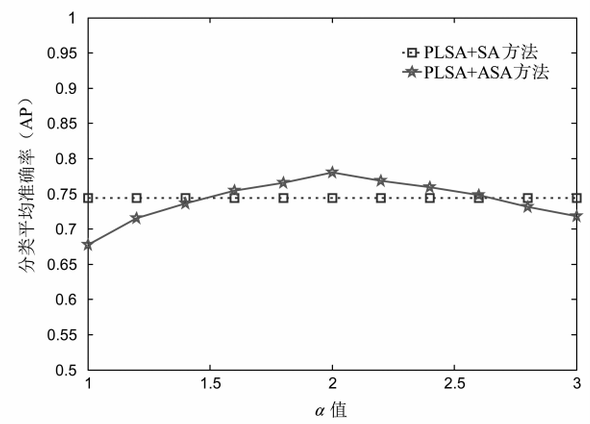

图5 自适应软分配因子α对AP值的影响

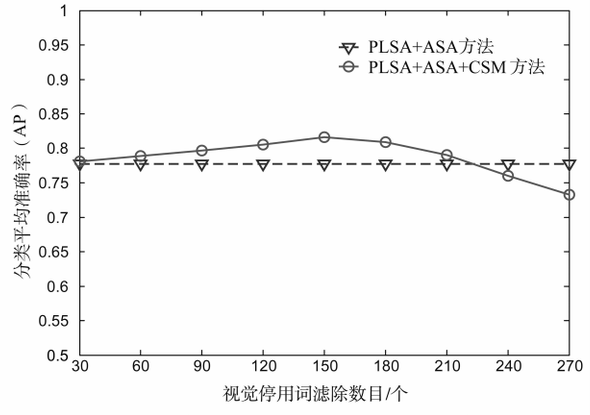

图6 视觉停用词滤除数目对目标分类AP值的影响

示. 从图6不难看出,采用卡方模型滤除一定数目的“视觉停用词”能够在一定程度上提高目标分类准确率,并且在滤除数目 $S=150$ 时能够达到最好的分类性能,即 AP 值为 81.53%. 然而,当滤除的单词数目过多时,难免使一些代表性强的单词也被错误地滤除,进而导致目标分类性能降低.

此外,图7给出了未进行视觉停用词滤除时,采用文中的近义词自适应软分配方法(PLSA + ASA)对随机选取的15类测试集进行目标分类结果的混淆矩阵图. 图8则给出了利用卡方模型滤除“视觉停用词”数目 $S=150$ 时对这15类测试集进行目标分类结果的混淆矩阵图. 从图7和图8中可以看出,采用本文方法(PLSA + ASA + CSM)进行目标分类时,多个目标分类的召回率均保持较高水平,且滤除“视觉停用词”可以使目标分类的召回率均有一定的提升. 但是,由于训练数据中各目标类别的差异性,所以针对不同的目标类别而言,滤除相同数目的“视觉停用词”对其性能改善程度略有不同,图7和图8能从另一个方面说明利用本文方法滤除视觉停用词对提高分类性能的有效性.

最后,为了进一步验证本文方法的有效性,又在 Pascal Voc2007 图像集[25]中的上进行实验,分别将 trainval 子集和 test 子集作为训练集和测试集,词典规模为 10K. 将本文方法(PLSA + ASA + CSM,文中参数分别为 $\alpha=2.4$, $m=20$, $S=1200$)与文献[8]中基于硬分配的视觉词典模型方法(HA)、文献[12]中基于软分配的视觉词典模型方法(SA)、文献[13]中基于上下文信息的视觉词典模型方法(Contextual-BoVW)以及文献[14]的基于 LDA 模型的软分配方法(LDA + SA)进行比较,得各目标分类准确率如表2所示. 从表2可以看出,SA 方法及 Contextual-BoVW 方法由于都引入一定的策略来克服视觉单词同义性和歧义性带来的量化误差严重等问题,其分类效果明显优于 HA 方法. 而 LDA + SA 方法在 SA 方法的基础上又利用 LDA 模型实现了更为准确的图像内容表达,因此,其分类准确率得到进一步改善. 而本文方法能够很好地从语义空间分析视觉单词间的远近,且采用了一种自适应软分配策略,并利用卡方模型滤除部分“视觉停用词”,因而较于其他方法能够取得最好的分类准确率.

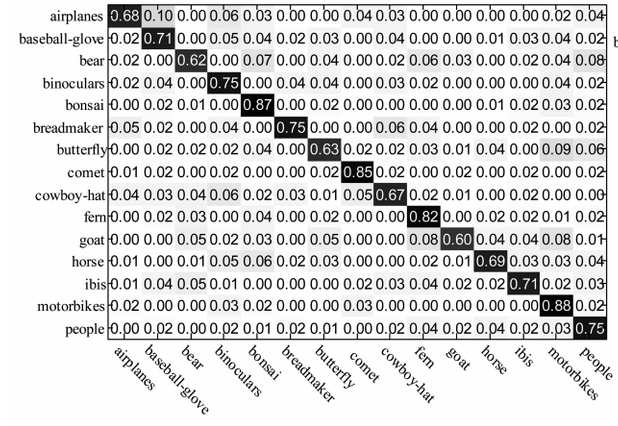

图7 未滤除“视觉停用词”时的目标分类结果的混淆矩阵

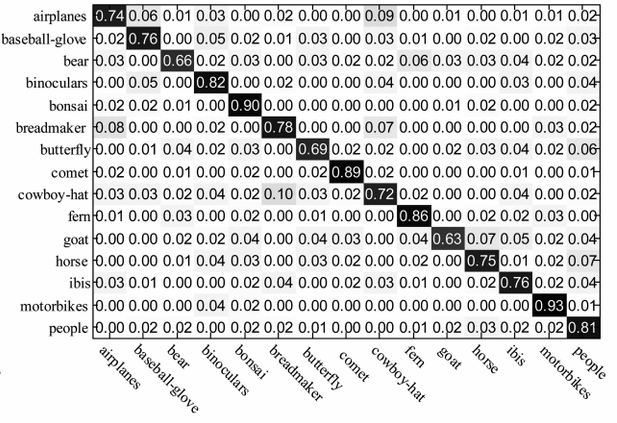

图8 “视觉停用词”滤除数目S=150时目标分类结果的混淆矩阵

表 2　不同方法在 Pascal Voc2007 图像集上的目标分类结果

目标类别	HA(%)	SA(%)	Contextual-BoVW(%)	LDA-SA(%)	PLSA + ASA + CSM(%)
airplanes	71.3	76.5	79.6	81.7	83.1
bicycle	67.1	73.8	77.1	79.5	82.7
bird	62.5	67.5	69.4	72.1	75.8
boat	66.7	73.1	78.2	79.5	82.4
bottle	46.7	55.7	63.1	66.4	68.5
bus	70.2	74.9	77.8	80.4	83.4
car	73.8	79.6	83.1	85.8	87.2
cat	62.7	68.6	73.6	76.4	76.3
chair	67.8	70.8	74.2	77.1	80.7
cow	68.1	74.3	77.6	80.4	85.1
diningtable	66.4	71.4	75.3	76.8	83.6
dog	54.5	64.5	69.1	74.2	79.4
horse	79.6	84.7	86.4	88.3	92.1
motorbike	70.6	75.0	77.6	78.5	80.5
person	85.9	90.1	91.6	91.4	93.4
pottedplant	58.7	65.0	72.8	75.4	79.3
sheep	62.4	68.1	73.2	76.1	80.2
sofa	61.9	68.2	71.6	73.2	77.6
train	82.6	89.5	92.4	92.6	94.5
Tvmonitor	61.4	66.6	70.3	73.4	76.7
Average	66.85	72.89	75.65	78.96	81.13

4　结语

本文首先采用概率潜在语义分析模型得到语义主题在某视觉单词下的概率分布,进而引入 K-L 散度度量视觉单词间的语义相关性,得到语义空间相近的近义词. 然后,根据图像各 SIFT 特征点模糊性类别自适应地完成特征点与若干近义词之间的软分配. 最后,采用卡方模型统计各视觉单词与图像类别的相关性,滤除“视觉停用词”,重构视觉词汇分布直方图,并由 SVM 分类器完成目标分类. 实验结果较好地验证了本文方法对克服视觉单词同义性和歧义性及量化误差问题的有效性,并能够有效地滤除视觉词典中的“视觉停用词”,进而提高目标分类性能. 需要指出的是,本文方法在语义层面分析视觉单词间的距离的同时,缺少有效的度量 SIFT 特征点与视觉单词间语义距离的方法,这在一定程度上会影响本文方法的性能. 因此,如何通过距离度量的学习使得特征空间的距离更加接近真实的语义距离是今后亟待解决的问题.

参考文献

[1] Sivic J, Zisserman A. Video Google: a text retrieval approach to object matching in videos[A] Proceedings of the 9th IEEE International Conference on Computer Vision [C]. Nice: IEEE Press, 2003. 1470-1477.

[2] 刘硕研,须德,冯松鹤,等. 一种基于上下文语义信息的图像块视觉单词生成算法[J]. 电子学报,2010,38(5):1156 – 1161.
LIU Shuo-yan, XU De, FENG Song-he, et al. A novel visual words definition algorithm of image patch based on contextual semantic information[J]. Acta Electronica Sinica, 2010, 38(5): 1156 – 1161. (in Chinese).

[3] 冯松鹤,郎丛妍,须德. 一种融合图学习和区域显著性分析的图像检索算法[J]. 电子学报,2011,39(10):2288 – 2294.
FENG Song-he, LANG Cong-yan, XU De. Combining graph learning and region saliency analysis for content-based image retrieval[J]. Acta Electronica Sinica, 2011, 39(10): 2288 – 2294. (in Chinese)

[4] Chen Y Z, Dick A, Li X, et al. Spatially aware feature selection and weighting for object retrieval[J]. Image and Vision Computing, 2013, 31(6): 935 – 948.

[5] Wang J Y, Bensmail H, Gao X. Joint learning and weighting of visual vocabulary for bag-of-feature based tissue classification[J]. Pattern Recognition, 2013, 46(3): 3249 – 3255.

[6] Otávio A B, Penatti, Fernanda B, et al. Visual word spatial arrangement for image retrieval and classification[J]. Pattern Recognition, 2014, 47 (1): 705 – 720.

[7] Nister D, Stewenius H. Scalable recognition with a vocabulary tree[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE Press, 2006. 2161 – 2168.

[8] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching[A]. Proceedings of IEEE Conference on Computer Vision and Pattern

Recognition[C]. Minneapolis:IEEE Press,2007. 1 – 8.
[9] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91 – 110.
[10] Van G J C,Veenman C J,Smeulders A W M,et al. Visual word ambiguity[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2010,7(32):1271 – 1283.
[11] Philbin J,Chum O,Isard M,et al. Lost in quantization: Improving particular object retrieval in large scale image databases[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Anchorage: IEEE Press,2009. 278 – 286.
[12] Koniusz P,Mikolajczyk K. Soft Assignment of visual words as linear coordinate coding and optimisation of its reconstruction error[A]. Proceedings of 18th IEEE International Conference on Image Processing[C]. Brussels: IEEE Press,2011. 2413 – 2416.
[13] Li T,Mei T,Kweon I S,et al. Contextual bags-of-words for visual categorization[J]. IEEE Transactions on Circuits System Video Technology,2012,21(4),381 – 392.
[14] Weinshall D,Levi G,Hanukaev D. LDA topic model with soft assignment of descriptors to words[A] Proceedings of the 30th International Conference on Machine Learning [C]. Atlanta:JMLR Press,2013. 711 – 719.
[15] Danilo D,Carneiro G,Chin T J,et al. Fuzzy clustering based encoding for Visual Object Classification[A]. Proceedings of IFSA World Congress and NAFIPS Annual Meeting[C]. Joint:IEEE Press,2013. 1439 – 1444.
[16] Su Y,Jurie F. Visual word disambiguation by semantic contexts[A]. Proceedings of International Conference on Computer Vision[C]. Barcelona:Springer,2011. 311 – 318.
[17] Liu S,Bai X. Discriminative features for image classification and retrieval[J]. Pattern Recognition Letters,2012, 33(6):744 – 751.
[18] Yuan J,Wu Y,Yang M. Discovery of collocation patterns:From visual words to visual phrases[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Rhode Island:IEEE Press,2012. 1 – 8.
[19] Chen T,Yap K H,Zhang D J. Discriminative soft bag-of-visual phrase for mobile landmark recognition[J]. IEEE Transactions on Multimedia,2014,16(3):612 – 622.
[20] Yeh J B,Wu C H. Extraction of robust visual phrases using graph mining for image retrieval[A]. Proceedings of IEEE Conference on Multimedia and Expo[C]. Singapore:IEEE Press,2010. 3681 – 3684.
[21] Hoffmann T. Probabilistic latent semantic analysis[A]. Proceedings of 15th Uncertainty in Artificial Intelligence [C]. Stockholm Sweden:AUAI Press,1999. 289 – 296.
[22] Emrah E,Nafiz A. Scene classification using spatial pyramid of latent topics[A]. Proceedings of IEEE 20th International Conference on Pattern Recognition[C]. San Francisco:IEEE Press,2010. 3603 – 3606.
[23] 张瑞杰,李弼程,魏福山. 基于多尺度上下文语义信息的图像场景分类算法[J]. 电子学报,2014,42(4):646 – 652.
Zhang Ruijie,Li Bicheng,Wei Fushan. Image scene classification based on multi-Scale and contextual semantic information[J]. Acta Electronica Sinica,2014,42(4):646 – 652. (in Chinese)
[24] Kesorn K,Poslad S. An enhanced bag-of-visual word vector space model to represent visual content in athletics images[J]. IEEE Transactions on Multimedia,2012,14(1): 211 – 222.
[25] Everingham M,Van Gool L,Williams C K I,et al. The PASCAL Visual Object Classes Challenge 2007 (VOC 2007) Results[DB/OL]. http://pascallin. ecs. soton. ac. uk/challenges/VOC/voc2007/results/index. shtml, 2014 – 05 – 11.
[26] Chang Chih Chung,Lin CJ. LIBSVM-A library for support vector machines[DB/OL]. http://www. csie. ntu. edu. tw/cjlin/libsvm/. 2014 – 04 – 12.

作者简介

赵永威 男,1988 年 1 月生于河南省周口市. 2015 年毕业于解放军信息工程大学获博士学位. 现为武警工程大学电子技术系讲师,主要研究方向为图像分析及处理.
E-mail:zhaoyongwei369@163. com

周 苑 女,1978 年出生,河南镇平县人,2006 年毕业于华中科技大学,获硕士学位,现为河南工程学院讲师,主要研究方向为媒体技术、计算机应用.
E-mail:363078125@qq. com

李弼程 男,1970 年生于湖南省衡阳市. 1998 年毕业于国防科技大学获博士学位. 现为解放军信息工程大学教授、博士生导师. 主要研究方向为智能信息处理.
E-mail:lbclm@163. com

柯圣财 男,1991 年生于湖北黄石市. 2013 毕业于解放军信息工程大学. 现为解放军信息工程大学硕士研究生,主要研究方向为图像检索.
E-mail:ke-shengcai@163. com