文章编号:1009-3087(2016)02-0169-07 DOI:10.15961/j.jsuese.2016.02.025

基于 E^2 LSH 过滤与空间一致性度量的目标检索方法

赵永威[1],李弼程[2],彭天强[3],唐永旺[2]
(1.武警工程大学 电子技术系,西安 陕西 710086;2.信息工程大学 信息系统工程学院,郑州 河南,450001
3.河南工程学院 计算机学院,郑州 河南,450001)

摘 要:为了解决传统视觉词典模型(bag of visual words model,BoVWM)中存在的时间效率低、词典区分性不强的问题,以及由于空间信息的缺失及量化误差等导致的目标检索性能较低的问题。提出一种新的目标检索方法,首先引入精确欧氏位置敏感哈希(exact euclidean locality sensitive hashing,E^2LSH)过滤训练图像集中的噪声和相似关键点,提高词典生成效率和质量;然后,引入卡方模型(Chi-square model)移除词典中的视觉停用词增强视觉词典的区分性;最后,采用空间一致性度量准则进行目标检索并对初始结果进行 *K*-近邻(*K*-nearest neighbors,*K*-NN)重排序。将提出的方法在数据库 Oxford5K 和 Flickr1 上进行目标检索,结果表明,新方法在一定程度上改善了视觉词典的质量,增强了视觉语义分辨能力,有效地提高目标检索性能。

关键词:目标检索;视觉词典模型;精确欧氏位置敏感哈希;空间一致性度量;卡方模型
中图分类号:TP391 **文献标志码**:A

Object Retrieval Based on E^2LSH Elimination and Spatially-constrained Similarity Measure

ZHAO Yongwei[1],*LI Bicheng*[2],*PENG Tianqiang*[3],*TANG Yongwang*[2]
(1. Dept. of Electrical Eng., Eng. Univ. of CAPF, Xi an, Xi'an 710086, China; 2. Inst. of Info. System Eng., Info. Eng. Univ., Zhengzhou 450002, China;
3. Dept. of Computer Sci. and Eng., Henan Inst. of Eng., Zhengzhou 450001, China)

Abstract: In order to resolve the problems of bag of visual words model (BoVWM) based object retrieval methods, such as low time efficiency, low distinction of visual words and weakly visual semantic resolution because of missing spatial information and quantization error, a novel object retrieval method was proposed. Firstly, E2LSH is used to identify and eliminate the noise key points and similar key points, consequently, the efficiency and quality of visual words was improved. Then, the stop words of dictionary were eliminated by Chi-square model to improve the distinguish ability of visual dictionary. Finally, the spatially-constrained similarity measurement was introduced to accomplish object retrieval, and a robust re-ranking method with the *K*-nearest neighbors of the query for automatically refining the initial search results was introduced. Experimental results on Oxford5K and Flickr1 datasets indicated that the distinguish ability of visual semantic expression is effectively improved and the object retrieval performance is substantially boosted compared with the traditional methods.

Key words: object retrieval; bag of visual words model; E^2LSH; spatially-constrained similarity measure; Chi-square model

近年来,随着图像数据规模的增大,使得图像处理面临的环境更加复杂。视觉词典模型(bag of visual words model,BoVWM)[1-3]由于其突出性能,已成为当前目标检索[4]、图像分类[5-6]等领域的主要解决方法。但是,以下几个关键性问题的存在极大地限制了 BoVWM 模型的性能。一是,关键点检测算子会产生大量的噪声点无疑会增加计算消耗,降低词典生成效率;二是,当前聚类算法的局限性[7-8]和图像背景噪声的存在,使得聚类生成的词典中包含一些类似于文本信息中的"的"、"和"、"是"等"停用词",这里称其为"视觉停用词",严重影响了视觉词典的质量;三是,传统的 BoVWM 模型中视觉单词间空间信息的缺失和量化误差严重等导致的视觉语义表达分辨力不强的问题。

收稿日期:2015-03-16
基金项目:国家自然科学基金资助项目(60872142;61301232);全军军事学研究生课题资助项目(YJS1062)
作者简介:赵永威(1988—),男,讲师,博士. 研究方向:图像分析及处理. E-mail:zhaoyongwei369@163.com

研究人员针对这些问题做了许多探索性研究,如在过滤噪声关键点方面:Rudinac 等[9]将相互距离小于1个像素值的特征点看作相似的近邻点,然后计算其中心值作为代表性特征点,这种方法最大的缺点是计算开销大,因为它需要遍历图像的每个像素点。Jamshy 等[10]通过学习特征点对某一特定应用的先验知识来过滤大部分特征点,然而这种方法却降低了图像分类性能。而针对"视觉停用词"去除问题,Sivic 等[2]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"停用词"过滤方法,然而,这种方法却忽略了视觉单词和目标语义概念间的相互关系。Tirilly 等[11]则根据关键点的几何性和概率隐语义分析模型淘汰无用的视觉单词,Yuan 等[12]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之间的空间关系。

针对视觉单词间空间信息的缺失和量化误差严重的问题,刘硕研等[13]采用一种基于上下文语义信息的图像块视觉单词生成算法,利用 PLSA 模型和 Markov 随机场共同挖掘单词的上下文信息。张瑞杰等[14]考虑到图像多尺度空间与单词上下文语义共生关系,在不同的图像尺度空间挖掘单词的上下文语义信息,进一步弥补了传统 BoVWM 模型的空间信息不足问题。Chen 等[8]则提出了一种基于软分配的视觉词组(visual phrase)构建方法,在弥补视觉单词空间信息的同时,有效克服了传统视觉词组构建方法[15]导致的特征信息丢失问题。而为了减小量化误差,van Gemert 等[16]提出了视觉单词不确定性(visual word uncertainty)模型,该模型同样是采用软分配策略对 SIFT 特征编码,进一步验证了软分配方法对于减弱视觉单词同义性和歧义性影响的有效性。Otávio 等[17]则提出一种基于视觉单词空间分布的图像检索和分类方法,该方法将视觉单词的空间信息嵌入到向量空间中,并对单词的在图像中的相对位置关系进行编码,从而得到更为紧致的视觉表达方式。Yang 等[18]则利用视觉语言模型结合目标区域周围的视觉单元构建了包含上下文语义信息的目标语言模型,进一步改善了目标检索性能。此外,文献[19]在利用上下文近义词构建视觉词汇直方图的同时,结合查询扩展方法解决目标视角变化较大、目标遮挡严重的情况问题。但是,查询扩展方法都依赖于较高的初始查全率,在初始查全率较低时,反而会带来一些负面影响。

针对上述问题,提出一种新的目标检索方法(enhanced visual dictionary and spatially-constrained similarity measure, EVD + SCSM)。首先,利用精确欧氏位置敏感哈希算法[20](exact euclidean locality sensitive hashing, E^2LSH)对位置的敏感性和处理高维数据的有效性,对图像初始关键点进行过滤,滤除噪声点的影响,降低计算消耗;然后,根据采用卡方模型分析视觉单词与目标类别的相关性大小,并按照从小到大的顺序滤除一定数量的视觉停用词,进一步优化视觉词典质量。最后,采用一种包含特征点角度、方向等空间一致性信息的度量方法完成目标检索,并引入 K-近邻重排序方法,进一步改善目标检索在复杂环境下的性能。

1 视觉词典优化

1.1 关键点过滤

假设在视觉位置相近的关键点是相似的,将其捆绑在一块,计算其质心,并将其作为一个有代表性的关键点。每个关键点 $p_i = \{\boldsymbol{u}_i, s_i, \theta_i, \boldsymbol{r}_i\}$ 由4部分组成,分别为:特征点在图像中的位置坐标 $\boldsymbol{u}_i$,特征的尺度 s_i,主方向 θ_i 及128维SIFT描述向量 $\boldsymbol{r}_i$。为了提高过滤效果,选取 $k(k=6)$ 个位置敏感函数联合起来以拉大关键点碰撞概率之间的差距,定义函数族:

$$G = \{g: S \rightarrow U^k\} \tag{1}$$

其中,$g(p) = \{h_1(p), \cdots, h_k(p)\}$。由式(1)可知,经函数 $g(p) \in G$ 降维映射后,关键点 p 都会变为一个 k 维向量 $\boldsymbol{a} = (a_1, a_2, \cdots a_k)$,然后,再采用主次哈希函数 h_1、h_2 对向量 $\boldsymbol{a}$ 进行哈希,构建哈希表并存储关键点。主次哈希函数的定义如下:

$$h_1(\boldsymbol{a}) = \left(\left(\sum_{i=1}^{k} r_i' a_i\right) \bmod m\right) \bmod z \tag{2}$$

$$h_2(\boldsymbol{a}) = \left(\sum_{i=1}^{k} r_i'' a_i\right) \bmod m \tag{3}$$

其中:r_i' 和 r_i'' 均为随机整数;z 为图像关键点总数目;m 为一个大的素数,通常取 $2^{32} - 5$。

由文献[6]的研究表明,X-means 算法是当前关键点过滤方法中较为有效的主流过滤方法,为此,分别采用 E^2LSH 和 X-means 算法对随机产生的数据点进行过滤以验证 E^2LSH 算法的有效性。如图1所示。由图1不难看出,X-means 方法过滤得到的代表性关键点较为不均,而由 E^2LSH 过滤得到的代表性关键点更为均匀。因此,从图1不难看出,基于 E^2LSH 的过滤方法能够提高过滤后各关键点的代表性和区分能力。

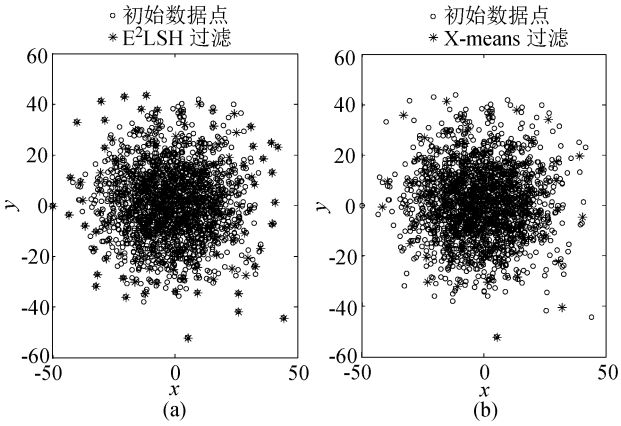

图 1　不同方法对关键点过滤的示意图

Fig. 1　Key points filtering diagram of different methods

1.2　“视觉停用词”去除

假设视觉单词 w 的出现频次独立于目标类别 C_j，$C_j \in C, 1 \le j \le k$，图像集 $C = \{C_1, C_2, \cdots, C_k\}$，那么，视觉单词 w 与各图像类别的卡方值可计算如下：

$$x^2 = \delta = \sum_{i=1}^{2}\sum_{j=1}^{k} \frac{(N \cdot n_{ij} - n_{i+} \cdot n_{+j})}{N \cdot n_{i+} \cdot n_{+j}} \tag{4}$$

其中，n_{1j} 为目标类别 C_j 包含单词 w 的图像数目，n_{2j} 为目标类别 C_j 不包含单词 w 的图像数目，n_{+j} 则为目标类别 C_j 中的图像总数，$n_{i+}(i = 1,2)$ 分别为图像集 C 中包含单词 w 的和不包含 w 的图像总数。

卡方值 x^2 代表了 w 与各目标类别间统计相关性的大小，同时考虑到单词 w 词频的影响，对卡方值赋予权重如下：

$$\tilde{x}^2 = \frac{x^2}{tf(w)} \tag{5}$$

其中，$tf(w)$ 为单词 w 词频。由此，就能够按照式(5)对每个单词的卡方值进行排序，然后去除一定数量 M 的“视觉停用词”即可。

2　相似性度量准则

2.1　空间一致性度量方法

假设一幅由矩形框界定好的查询目标图像，其空间信息可以表示如下：$B = \{x_c, y_c, v, h, \theta\}$，如图 2($a$) 所示，其中，$(x_c, y_c)$ 为界定目标的矩形框中心坐标，v、h 分别为矩形框的宽和高，θ 为矩形框的旋转角度。通过相似变换 T 就能检索到图像库中与之最匹配的图像，$T = \{\boldsymbol{R}(a), s, \boldsymbol{t}\}$，$a$ 为目标的旋转角度，$\boldsymbol{R}(a) = \begin{bmatrix} \cos a & \sin a \\ \sin a & \cos a \end{bmatrix}$，$s$ 为尺度变化，$\boldsymbol{t} = (x_t, y_t)$ 为位置变化，那么经过变换后的目标图像即为：$B' = T(B) = \{x_c + x_t, y_c + y_t, s \cdot v, s \cdot h, \theta = a\}$。这里，用 Q 表示查询图像，D 表示图像库中任一幅图像，其 SIFT 特征点分别表示为 $\{f_1, f_2, \cdots, f_m\}$、$\{g_1, g_2, \cdots, g_n\}$，那么，2 幅图像之间的空间一致性度量可计算如下：

$$S(Q, D \mid T) = \sum_{k=1}^{n} \frac{idf^2(w_k)}{tf_Q(w_k) \cdot tf_D(w_k)},$$
$$\text{s.t.} \begin{cases} w(f_i) = w(g_j) = w_k, \\ \| T(\boldsymbol{L}(f_i) - \boldsymbol{L}(g_j)) \| < \varepsilon \end{cases} \tag{6}$$

其中：w_k 为视觉词典里的第 k 个单词；n 为词典规模；$w(f_i) = w(g_j) = w_k$ 表示特征点 f_i、g_j 都被映射至单词 w_k 上；$\boldsymbol{L}(f) = (x_f, y_f)$ 表示特征点的位置；$\| T(\boldsymbol{L}(f_i) - \boldsymbol{L}(g_j)) \| < \varepsilon$ 为 2 个相互匹配的特征点之间的约束条件，保证它们经过变换之后位置依然较近；$idf(w_k)$ 为单词 w_k 的逆文档频率；$tf_Q(w_k)$ 为单词 w_k 在图像 Q 中的词频；$tf_D(w_k)$ 为在图像 D 中的词频。因此，对于图像库中的每一幅图像而言，就是要找到最优变换 T^* 使得度量值最大。即

$$T^* = \{\boldsymbol{R}(a^*), s^*, \boldsymbol{t}^*\} = \arg\max_T S(Q, D \mid T) \tag{7}$$

故而，就有 $S^*(Q, D) = S(Q, D \mid T^*)$ 可以用来衡量图像 Q 和图像 D 之间的相似性，且所有的检索结果也能以此进行排序。由图 2(a)、(b)不难看出，2 幅图像中只有特征点 (f_i, g_i)，$i = 1,2,3$ 是满足空间一致性条件的。(f_5, g_5) 是一个错误匹配点对，(f_4, g_4) 的取舍则决定于式(6)中参数 ε 的大小。

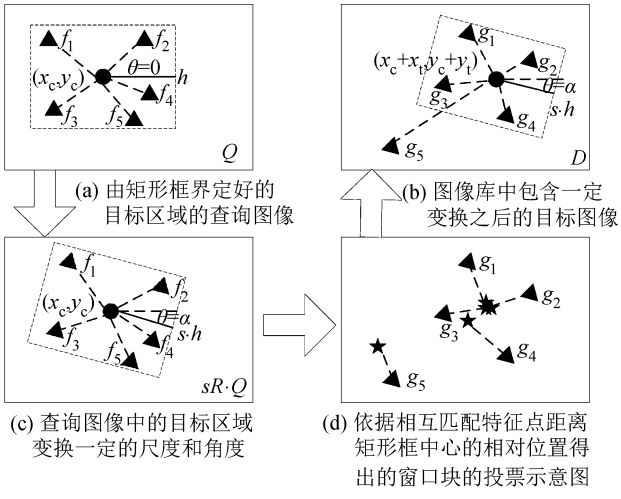

图 2　空间一致性度量示意图

Fig. 2　Illustration of spatially-constrained similarity measurement

为了计算 $S^*(Q, D)$，需要找到最优变换 T^*，这里可将对 T 进行分解处理，首先将 360° 空间划分 n_R 部分(一般 $n_R = 4$ 或 8)，同样地，尺度空间被划分为 n_S 部分，通常 $n_S = 8$，变化范围为 1/2 ~ 2。令 $\boldsymbol{V}(f)$ 表示特征点 f 与查询图像中的矩形框中心 c_Q 之间的

相对位置关系向量,那么由匹配的特征点对(f,g)的位置及$\boldsymbol{V}(f)$就能定位图像D中的矩形框中心,$\boldsymbol{L}(c_Q)=\boldsymbol{L}(g)-\boldsymbol{V}(f)$,如果$w(f)=w(g)=w_k$,特征点对$(f,g)$的投票得分为:

$$Score(w_k)=\frac{idf^2(w_k)}{tf_Q(w_k)\cdot tf_D(w_k)} \tag{8}$$

不难看出,若相互匹配的特征点对符合空间一致性条件,那么由其投票得出的矩形框中心位置也是相近的,如图2(d)所示。每次投票得出的目标位置中心就代表了一个变换T,那么利用式(8)投票所得分数就等同于利用式(6)进行相似性度量。在实际应用中,可将投票得分图归一化为$n_x\times n_y$个图像块,同时为了避免投票时的量化误差及弱化目标遮挡等情况的影响,对所估计的中心块周围16×16像素的窗口块进行投票,而每个块的得分大小为$Score(w_k)\times e^{-d/\sigma^2}$,其中,$e^{-d/\sigma^2}$为权重系数,由每个块与中心块之间的距离$d$和参数$\sigma$决定,整个过程相当于对投票得分图进行一次高斯平滑。

2.2 *K*-近邻重排序

检索结果记为$R(Q,D)$,并令N_i表示查询图像的第i个检索结果,则有$R(Q,N_i)=i$,用$N_q=\{N_i\}$,$i=1,2,\cdots,k$表示查询图像的K-近邻。为了有效地利用K-近邻图像包含的信息,本文重新利用其中的每一幅图像作为查询图像重新检索,并分别将排序结果记为$R(N_i,D)$,依据这个排序结果给图像库中的每幅图像分配一个得分$\frac{1}{R(N_i,D)}$,那么经重排序之后的图像之间的一致性度量可定义为:

$$\bar{S}(Q,D)=\frac{\omega_0}{R(Q,D)}+\sum_{i=1}^{k}\frac{\omega_i}{R(N_i,D)} \tag{9}$$

其中,ω_i为权重系数,由初始排序决定。这里,令$\omega_0=1$,$\omega_i=\frac{1}{1+R(N_i,D)}=\frac{1}{1+i}$,若将查询图像本身看作其第0近邻,那么式(9)可转化为:

$$\bar{S}(Q,D)=\sum_{i=0}^{k}\frac{\omega_i}{R(N_i,D)}=\frac{1}{(i+1)R(N_i,D)} \tag{10}$$

其中,$\bar{S}(Q,D)$为一个单向性度量,而只有$R(Q,N_i)$和$R(N_i,Q)$同时排在前列才能保证两者互为近邻。为此,可将权重系数ω_i改为:

$$\omega_i=\frac{1}{(R(Q,N_i)+R(N_i,Q)+1)}=\frac{1}{(i+R(N_i,Q)+1)},$$

由此可以得到最终的相似性度量准则为:

$$\bar{S}(Q,D)=\sum_{i=0}^{k}\frac{1}{(i+R(N_i,Q)+1)R(N_i,D)} \tag{11}$$

然后,所有图像即可按照式(11)进行重排序,完成检索。

3 实验设置与性能分析

3.1 实验设置

选取Oxford5K数据库[21]作为实验数据库,并从每个目标类别中选取50幅图像,共550幅图像作为训练图像库来生成视觉词典,词典规模为10×10^4。此外,引入Flickr1数据库[22]作为干扰数据以验证提出的方法在复杂环境下的实验性能。实验硬件配置为Core 2.6 GHz×4,内存4 GB的台式机,软件环境为MATLAB2012a,性能评价采用*AP*和*MAP*值以及时间效率。其中,*AP*(average precision)为查准率-查全率曲线所包含的面积,而*MAP*为5幅查询图像的平均*AP*值。相关定义如下:

$$\text{查全率}=\frac{\text{检索出的相关图像}}{\text{全部相关图像}}\times100\% \tag{12}$$

$$\text{查准率}=\frac{\text{检索出的相关图像}}{\text{检索出的全部图像}}\times100\% \tag{13}$$

3.2 实验性能分析

实验从550幅训练图像库中提取约1 436 634个特征点,利用E^{2}LSH对其过滤,并采用AKM聚类算法对未过滤关键点和不同k值过滤后的特征点进行聚类,生成相同单词数目的词典进行目标检索分析了参数k对目标检索结果*MAP*值的影响(此时,令$\sigma^2=0$),如图3所示。

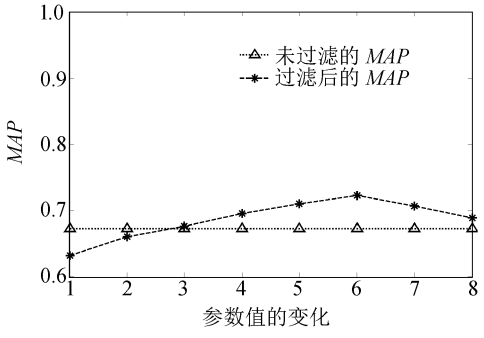

图3 参数k对目标检索*MAP*值的影响

Fig. 3 Influence of parameter k for object retrieval *MAP* results

从图3中不难看出,随着参数k值的变化,目标检索的*MAP*也随之变化,且在$k>3$时,经E^{2}LSH过滤后的检索*MAP*值要高于未过滤的目标检索。且当$k=6$时,目标检索*MAP*值最大,这是因为,当k值较小时会使得过滤后的关键点数目过少,从而容易丢失图像包含的细节信息,而当k值过大时导致过滤后的特征点数目过多,使得算法过滤效果不明

显,综合考虑,取 $k = 6$, 此时剩余代表性关键点数目为 1 002 105 个,过滤率为 31.3%。实验又将提出的方法与传统的 AKM 算法在生成视觉词典时的时间消耗作了对比,具体如图 4 所示。从图 4 中可以看出,经算法的过滤以后,视觉词典的生成效率有较为明显提升。

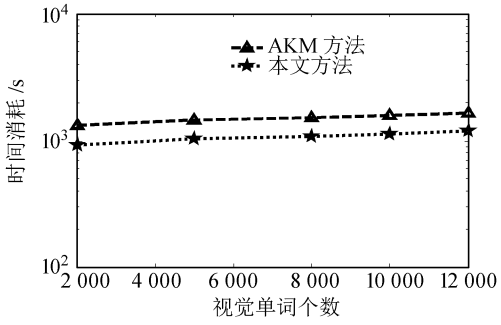

图 4　不同方法构建词典效率对比

Fig. 4　Efficiency comparison of different methods in constructing visual dictionary

实验在 E^2LSH 函数个数 $k = 6$ 的情况下,对关键点进行过滤,并生成规模为 10×10^4 的视觉词典,然后利用卡方模型滤除一定数量 M 的视觉停用词,验证过滤不同数目“视觉停用词”对目标检索结果的影响,并与未进行视觉停用词滤除时的目标检索结果进行对比,得其检索 *MAP* 值如图 5 所示。从图 5 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检索的 *MAP* 值,并且在滤除数目 $M = 1\,000$ 时能够达到最高的 *MAP* 值,即为 76.4%。同时,从图 5 中可以看出,当滤除的单词数目过多时,会导致目标检索性能降低,这是因为滤除过多的视觉单词难免使一些代表性强的单词也被错误地滤除。

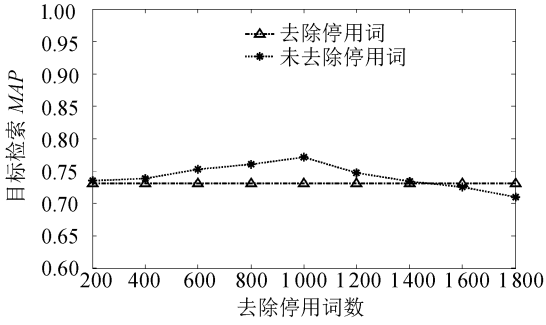

图 5　去除停用词数目对目标检索 *MAP* 值的影响

Fig. 5　Influence of parameter *M* for *MAP*

在 E^2LSH 函数个数 $k = 6$,去除视觉停用词数目 $M = 1\,000$ 的情况下,实验以 Oxford5K 为实验数据库分析了空间一致性度量准则中参数 σ^2 对目标检索 *MAP* 值的影响,结果如图 6 所示。其中,当 $\sigma^2 = 0$ 时表示不对投票结果进行高斯平滑,即每个匹配特征对都将票数投向根据式(8)所估计的一个中心块,由图 6 不难看出,当 $\sigma^2 > 0$ 时,也表示对所估计的中心块周围 16 × 16 窗口块进行投票的 *MAP* 值明显优于未对投票结果进行高斯平滑的情况(即 $\sigma^2 = 0$),且在 $\sigma^2 = 2.5$ 时取得最大的 *MAP* 值,因此,取 $\sigma^2 = 2.5$。

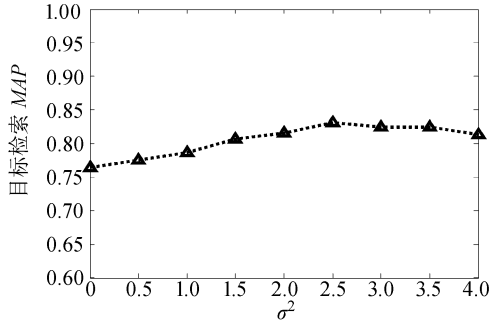

图 6　参数 σ^2 对 *MAP* 值的影响

Fig. 6　Influence of parameter σ^2 for *MAP*

为了验证提出的方法中空间一致性度量准则以及重排序方法对改善目标检索结果的有效性,实验将提出的方法(EVD + SCSM)与 AKM + LM (AKM + language model)方法[18]、CSVW + QE(contextual synonymous visual words + query expansion)方法[19]以及将优化的视觉词典与语言模型相结合的方法(enhanced visual dictionary + language model, EVD + LM)在 Oxford5K + Flickr1 数据库上对 11 个查询目标的检索准确度作了比较,得平均查准率均值(*MAP*)如表 1 所示。

表 1　不同方法的目标检索 *MAP* 值对比

Tab. 1　Object retrieval *MAP* results of different methods for Oxford5K database

查询目标	目标检索 *MAP* 值/%			
	AKM + LM	CSVW + QE	EVD + LM	EVD + SCSM
Ashmolean	53.6	72.6	69.2	81.9
All Souls	51.8	70.1	66.7	83.8
Balliol	54.5	66.2	64.9	79.5
Bodleian	47.9	59.8	57.3	69.4
Cornmarket	41.3	70.2	71.8	82.7
Christ Church	50.4	63.4	67.6	71.8
Magdalen	23.6	33.7	34.5	49.9
Hertford	75.7	81.0	80.2	91.0
Pitt Rivers	91.0	93.4	91.3	95.6
Keble	75.7	85.8	82.8	87.4
Radcliffe Cam	53.9	72.3	70.7	82.4
Average	56.30	69.86	68.81	79.58

从表 1 可知,对不同的查询目标而言,采用

AKM + LM 方法的 *MAP* 值均低于其他几种方法。而 EVD + LM 方法的 *MAP* 值相较于 AKM + LM 方法有一定的改善,足以说明提出的词典优化方法能有效降低图像背景噪声点和停用词的影响,提高视觉词典的区分性;同时 CSVW + QE 方法的性能要略好于 EVD + LM 方法,这是因为 CSVW + QE 方法在利用空间信息的基础上又结合查询扩展策略。得到了更多与查询目标相关的图像。但是,EVD + SCSM 方法的检索 *MAP* 值要远高于上述3种方法,与 EVD + LM 方法对比可以看出,提出的空间一致性度量准则对单词空间信息的利用优于视觉语言模型。

图7给出了 *K*-近邻重排序方法的效果示例图。从图7中可以看出,第4幅最近邻图像与查询目标图像无关,但是由其检索得到的虚线框中的任何一幅图像的最终检索得分不会改变,因为它们与其他的最近邻图像不相关。而用实线框标志的图像会得到较高的检索得分,因为它们与 *K*-近邻中的多数图像相关。不难看出,采用 *K*-近邻重排序方法之后可以得到更多与包含查询目标的图像。

图7 *K*-近邻重排序结果示意图

Fig. 7 Illustration of *K*-NN reranking results

4 结 论

为了改善生成视觉词典的质量,提高视觉单词对图像内容的表达能力,首先利用 E^2LSH 算法对图像初始关键点进行过滤,降低噪声点的影响;然后,引入卡方模型统计各视觉单词与目标类别的相关性,并结合单词词频信息移除词典中的视觉停用词;最后,为了确保度量的准确性,采用空间一致性度量准则进行相似性度量以弥补传统视觉词典模型中单词空间关系缺失降低量化误差并对初始检索结果进行 *K*-近邻重排序。实验结果有效地验证了提出的方法对改善目标检索性能的有效性。

需要注意的是,在今后需要研究如何降低 E^2LSH 算法的随机性问题来提高过滤效果的鲁棒性。此外,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也是今后亟待解决的问题。

参考文献:

[1] LoweD G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91 – 110.

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

[3] Jégou H, Douze M, Schmid C. Improving bag-of-features for large scale image search[J] Computer Vision, 2010, 87(3): 316 – 336.

[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(12): 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(12): 3249 – 3255.

[6] Cao Y, Wang C H, Li Z W, et al. Spatial-bag-of-features [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010: 3352 – 3359.

[7] Dai L, Song X Y, Wang F, et al. Large scale image retrieval with visual groups[C]//Proceedings of IEEE Conference on Image Processing. Melbourne, Australia: IEEE, 2013: 2582 – 2586.

[8] 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.

[9] Rudinac M, Lenseigne B, Jonker P. Keypoint extraction and selection for object recognition[C]//Proceedings of IEEE Conference on Machine Vision Applications. Sassari, Italy: IEEE, 2009: 191 – 194.

[10] Jamshy S, Krupka E, Yeshurun Y. Reducing keypoint database size[C]//Proceedings of 15th International Conference on Image Analysis and Processing. Vietrisul

Mare, Italy: IEEE, 2009: 113 – 122.

[11] Tirilly P, Claveau V, Gros P. Language modeling for bag of visual words image categorization[C]//Proceedings of 2008 International Conference on Content-Based Image and Video Retrieval. New York: IEEE, 2008: 249 – 258.

[12] Yuan J, Wu Y, Yang M. Discovery of collocation patterns: From visual words to visual phrases[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007: 1 – 8.

[13] Liu Shuoyan, Xu De, Feng Songhe, 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. [刘硕研,须德,冯松鹤,等. 一种基于上下文语义信息的图像视觉单词生成算法[J]. 电子学报, 2010, 38(5): 1156 – 1161.]

[14] Zhang Ruijie, Li Bicheng, Wei Fushan. Imagescene classification based on multi-scale and contextual semantic information[J]. Acta Electronica Sinica, 2014, 42(4): 646 – 652. [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.]

[15] Yeh J B, Wu C H. Extraction of robust visual phrases using graph mining for image retrieval[C]//Proceedings of IEEE Conference on Multimedia and Expo. Paris: IEEE, 2010: 3681 – 3684.

[16] 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 – 1283.

[17] Penatti O A B, Silva F B, Valle E, et al. Visual word spatial arrangement for image retrieval and classification[J]. Pattern Recognition, 2014, 47(2): 705 – 720.

[18] Yang L, Geng B, Cai Y, et al. Object retrieval using visual query context[J]. IEEE Transactions on Multimedia, 2012, 13(6): 1295 – 1307.

[19] Xie Hongtao, Zhang Yongdong, Tao Jianlong, et al. Contextual query expansion for image retrieval[J]. IEEE Transactions on Multimedia, 2014, 16(4): 1104 – 1114.

[20] Slaney M, Casey M. Locality-sensitive hashing for finding nearest neighbors[J]. IEEE Signal Processing Magazine, 2008, 8(3): 128 – 131.

[21] Robotics Research Group. Oxford5K dataset[DB/OL]. [2014 – 03]. http://www.robots.ox.ac.uk/_vgg/data/oxbuildings/.

[22] Yahoo Company. Flickr1 dataset[DB/OL]. [2014 – 03]. http://www.flickr.com/.

(编辑　赵　婧)