大规模图像集中的代表性图像选取

齐美彬 [1]　朱俊俊 [1]　纪 平 [2]　蒋建国 [1]

摘 要 针对传统图像检索系统通过关键字搜索图像时缺乏语义主题多样性的问题, 提出了一种基于互近邻一致性和近邻传播的代表性图像选取算法, 为每个查询选取与其相关的不同语义主题的图像集合. 该算法利用互近邻一致性调整图像间的相似度, 再进行近邻传播 (AP) 聚类将图像集分为若干簇, 最后通过簇排序选取代表性图像簇并从中选取中心图像为代表性图像. 实验表明, 本文方法的性能超过基于 K-means 的方法和基于 Greedy K-means 的方法, 所选图像能直观有效地概括源图像集的内容, 并且在语义上多样化.

关键词 代表性图像, 语义主题, 互近邻一致性, AP 聚类, 图像簇排名

引用格式 齐美彬, 朱俊俊, 纪平, 蒋建国. 大规模图像集中的代表性图像选取. 自动化学报, 2014, **40**(4): 706−712

DOI 10.3724/SP.J.1004.2014.00706

Representative Image Selection from Image Dataset

QI Mei-Bin[1]　ZHU Jun-Jun[1]　JI Ping[2]　JIANG Jian-Guo[1]

Abstract In a traditional image retrieval system, people search images using keywords. However, the result shows a lack of diversity in the sense of semantic theme. For the problem, we propose a viable method for representative image selection. We define representative images as those with diverse contents in the semantic meaning to cover different semantic forms of a query. First, we use mutual nearest neighbor consistency to adjust the similarity between images as the input to the AP clustering. Then we select representative clusters based on cluster ranking and finally take the images of the cluster center from representative clusters as a summary of the image dataset. The results showed that the performance of our method is better than the K-means based method and the greedy K-means based method. The selected images can summarize the content of the original image dataset intuitively and effectively, and they are diverse in semantic meaning as well.

Key words Representative images, semantic theme, AP clustering, mutual nearest neighbor consistency, cluster ranking

Citation Qi Mei-Bin, Zhu Jun-Jun, Ji Ping, Jiang Jian-Guo. Representative image selection from image dataset. *Acta Automatica Sinica*, 2014, **40**(4): 706−712

在图像分享网站中, 用户通过输入关键词查找图像, 返回的图像列表中大多数图像在视觉上相似, 缺少语义主题的多样性[1]. 例如在 Google 图像搜索引擎中将 tiger 作为查询词, 返回的前 21 个最相关的结果如图 1 所示. 图中可以看出查询结果在内容上很相似, 缺乏语义多样性. 然而很多时候, 用户事先并不确定自己真正感兴趣的内容, 因此如何获取用户意图[2−3] 是当前的一个热门话题. 例如当用户感兴趣的是图 2 中所示的 Tiger Woods 球星或者是 tiger 航班时, 在图 1 所示的返回结果中没有相关的内容. 因此, 创建一种包含了关于某个查询的多种语义主题的图像摘要[4](即代表性图像集合) 是很有必要的, 而且图像摘要技术可应用于旅游指南[5]、图片浏览、视频摘要[6] 及图像排序[7−8] 等多个领域.

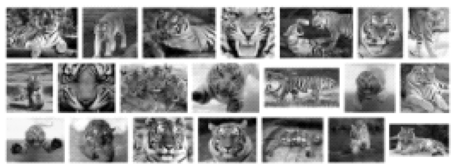


图 1　Google 图片搜索引擎查询 “tiger” 返回的前 21 个最相关的结果

Fig. 1　Top 21 “most relevant” images from Google for the query “tiger”

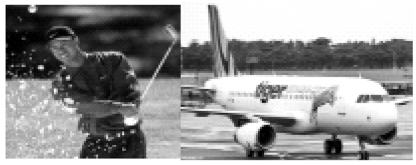


图 2　与查询 “tiger” 相关的图片

Fig. 2　Images relevant to query “tiger”

收稿日期 2012-11-26　录用日期 2013-03-04
Manuscript received November 26, 2012; accepted March 4, 2013
国家自然科学基金 (61172164) 资助
Supported by National Natural Science Foundation of China (61172164)
本文责任编委 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 合肥工业大学计算机与信息学院 合肥 230009　2. 合肥学院电子信息与工程系 合肥 230022
1. School of Computer and Information, Hefei University of Technology, Hefei 230009　2. Department of Electronics & Electrical Engineering, Hefei University, Hefei 230022

当前, 针对从网络相册集中选取最具代表性或最典型的图像问题, 研究人员已经开展了一系列的研究工作. 其中, 文献 [9] 基于文本标签从一组大规模附有地理位置标签的图像集中选取了一部分图像作为摘要子集. 通过对文本标签和地理位置标签对应关系的分析, 他们开发了基于地图的可视化工具 "Tag map", 用于显示最重要的地区和该地区最具代表性的图像. 文献 [10] 提出了一种基于图像聚类的方法, 对标志性建筑景观选出其典型的图像, 这种聚类方法主要利用了图像之间的视觉相似性. 聚类之后定义了一个图像质量指标, 根据质量指标来选择代表性图像. 与此类似, 文献 [11] 主要基于图像的聚类以及由图像相似特征点建立的连接结构, 为世界上著名的标志性建筑景观生成代表性的图像. 与他们的工作相比, 我们选择了更具有实际意义的一般事物 (包括目标、卡通、场景等) 作为研究对象, 而一般事物代表性图像的主观性和不确定性致使研究更加困难. 文献 [4] 利用核函数计算不同特征空间中图像的相似度, 并通过核学习和用户反馈调整各相似度之间的融合权值, 并在特征空间中选择距离聚类中心最近的图像作为本类别的代表性图像. 然而核函数的构造比较复杂, 很难找到真正合适的核函数, 能够精确描述两幅图像之间的相似度. 与本文研究类似的是文献 [12], 旨在为一般抽象化的语义类别, 如 "love"、"beauty"、"close up" 等来生成代表性的形象化标志图像. 目的在于为每一个类别选出一小部分典型的图像. 其方法将视觉特征聚类和标签聚类相结合, 并从审美和图像质量的角度选择代表性图像. 与其不同的是, 本文没有涉及对图像标签进行语义分析, 因此对图像聚类的准确度要求更高.

现有的方法主要采用基于聚类的方法来选取代表性图像. 首先, 利用一组高维的特征向量 (例如颜色、纹理及局部特征点等) 来表示图像集中的图像; 其次, 基于特征空间采用聚类算法将图像划分为若干图像簇; 最后, 对这些图像簇进行质量排名, 并从中选取经典图像[13−14] 作为图像摘要. 现有方法中常用的聚类算法是简单快速的 K-means 算法[15]. 然而 K-means 算法必须提前设定聚类的个数 K, 并且对初始值非常敏感. 对噪音和孤立点也同样敏感. 即使很少一部分的孤立点也将会对均值产生很大影响. 现有方法的另外一个不足之处是相似性度量. 例如余弦相似度经常用于相似性度量, 然而直接引用并不是十分有效, 不能准确地描述图像之间的相似程度.

为解决上述问题, 本文提出一种基于互近邻一致性和近邻传播 (AP) 的代表性图像选取算法[16]. 利用互近邻一致性[17] 对 AP 聚类算法中最重要的参数之一— 相似度矩阵 s 进行调整, 使得相似度矩阵 s 更能准确地描述图像之间的相似关系, 聚类的效果得到改善. 其次, 对聚类后得到的若干图像簇依据以下准则进行簇排名: 1) 视觉一致性. 同一个簇中的图像之间彼此应该很相似, 呈现视觉上的一致性, 而不同簇中的图像不相似; 2) 簇连通性. 如果一个簇中的图像与该簇中的其他图像都有连接, 则说明该图像簇具有代表性. 因为这些连接可能暗示在很多图像中都出现了某个相似的场景或目标. 选取排名靠前的图像簇作为代表性簇, 最后从中选取代表性图像. 实验结果证明了该方法的有效性.

本文的算法框架如图 3 所示, 包含两个部分: 特征提取和代表性图像选取. 其中, 特征提取分为两个步骤, 首先提取图像显著区域, 然后对显著区域提取 3 种特征, 包括颜色特征、纹理特征和局部点特征. 在代表性图像选取部分, 首先基于特征空间采用 AP 聚类算法将图像聚成若干图像簇, 然后进行簇排名, 并从中选取代表性图像.

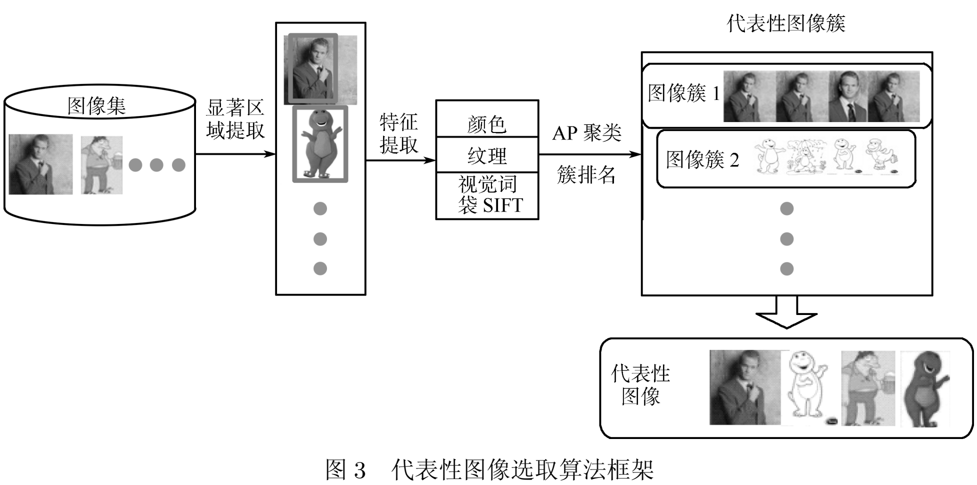


图 3　代表性图像选取算法框架

Fig. 3　The proposed framework of representative image selection arithmetic

1 特征提取和相似性度量

由于图像的显著区域能更简洁、准确地代表图像所表达的内容, 同时能够有效地削弱背景的干扰, 因此本文先获取图像的显著区域. 为此, 首先计算图像的显著图[18], 而后对显著图进行阈值分割, 过程如下:

1) 找出显著图中大于 T_1 的像素点 P_i, $i = 1, 2, \cdots, n$. 为了尽可能地分割出完整的目标, 将阈值 T_1 设置为较小的值, 取经验值 $T_1 = 5$;

2) 计算这 n 个点的平均值 $T^{average}$;

3) 为了去除某些背景的干扰, 保留这 n 个点中的灰度值大于某个阈值的点. 如果这个阈值取得过大, 则目标会因分割得过碎而出现多目标; 取值太小, 则会将过多的背景包含进来. 多次实验证明, 当阈值为 $4 \times T^{average}/3$ 时, 效果较好.

然后对产生的各个显著区域求取能恰好覆盖保留点的最小外接矩形, 即为图像的显著区域.

本文提取了图像的全局特征和局部特征. 其中全局特征包括 36 维的分块颜色累加直方图特征和 48 维的 Gabor 小波纹理特征; 局部特征是基于视觉词袋[19] 的 SIFT 特征[20−21], 维数为 1 024, 并采用软权重的思想, 克服了硬分配统一权重的问题. 图像相似度矩阵的定义直接影响基于相似度矩阵的聚类算法性能. 如果相似度矩阵能够准确地描述图像之间的相似关系, 则下一步进行 AP 聚类算法就能够得到鲁棒的聚类效果. 对图像提取了上述 3 种特征之后, 首先分别在各特征空间中计算图像的相似度, 本文采用较为常用的余弦相似度. 然后利用互近邻一致性对图像相似度进行调整, 即 K 互近邻点彼此的相似度值为 1, 这样保证 K 互近邻点 (图像) 在聚类后存在于同一个图像簇中. 由于 K 的取值直接影响聚类的效果, 且对于不同的图像集 K 的最优取值也是不一样的, 所以本文在不同的图像集上进行了实验, 最终选取适当的 K 值. 最后对各特征空间计算出的相似度进行融合, 通过大量的实验找出最优权重以适应不同的图像集要求, 得到融合了多特征的图像相似度矩阵.

2 选择代表性图像

2.1 基于 AP 算法的图像聚类

代表性图像即摘要图像, 是一组能够代表整个源图像集的内容梗概的图像, 用于方便用户进行浏览和查询. 因此摘要图像的选取问题通常转换为如何先将目标图像集聚合成一系列高质量的图像簇, 进而从中选取那些最具视觉内容代表性的图像作为整个图像的摘要问题.

AP 聚类算法是发表于《*Science*》上的一种新的基于相似度空间的聚类算法, 它可以在相似度空间中将目标图像集聚合成一组规模适度、内容紧凑的图像簇. 一般聚类中心往往被看作本类别的代表, 然而基于特征空间的聚类方法, 其聚类中心由特征向量加权迭代而生成, 并不是真实对应某个图像. 而 AP 的聚类算法与一般的聚类算法不同, 它的聚类中心对应数据集中真实存在的数据, 并且聚类中心图像具有很好的代表性. 因此我们采用 AP 聚类算法, 并将经过互近邻一致性调整后的图像相似度矩阵 s 作为其输入, 对图像集进行聚类.

2.2 生成代表性图像

经过 AP 聚类算法, 图像被分成了若干簇. 对于这些图像簇, 先根据它们的代表性程度进行排名. 然后从排名靠前的图像簇中选择代表性图像并返回给用户, 而排名较低的图像簇则被丢弃. 本文假设代表性图像簇具有以下特性:

1) 视觉一致性. 首先利用视觉特征来度量一个图像簇的簇内距离, 我们用 $intra\_dist$ 表示, 即该簇中的两两图像之间的平均距离. 然后计算簇间距离, 用 $inter\_dist$ 表示, 即该簇中的图像与簇外图像的平均距离. 最后计算簇间距离与簇内距离之比. 比值越大则说明该类中的图像越相似, 呈现视觉上的一致性; 比值越低则表明该图像簇中可能有噪音, 视觉上可能不一致, 或者该簇中的图像与其他簇中图像相似;

2) 簇连通性. 可以利用 SIFT 特征建立不同图像之间的连接, 我们采用图像簇中两两图像之间 SIFT 特征点匹配个数的平均数 $link\_num$ 来表示. 如果一个簇中的图像与该簇中的其他图像都有连接, 则说明该图像簇有较强的代表性. 因为这些连接可能暗示在很多图像中都出现了某个类似的场景或目标.

根据上述两种特性选出符合条件的图像簇, 作为最终的代表性图像簇, 并选择簇中心图像作为代表性图像. 因此, 对于每个图像簇, 计算过程如下:

1) 分别计算图像簇的簇内平均距离 $intra\_dist$ 和簇间距离 $inter\_dist$;

2) 计算簇间距离与簇内距离之比:

$$ratio = inter\_dist/intra\_dist \tag{1}$$

3) 计算簇内图像间的 SIFT 匹配点数的平均值 $link\_num$;

4) 对于所有图像簇的 $ratio$ 和 $link\_num$ 进行降序排序, 取前 n 个 $ratio$ 与前 n 个 $link\_num$ 所对应的图像簇的交集, 并将这些交集图像簇作为最终的代表性图像簇. 在本文的实验中, n 的取值为 5. 一般来说, 对于某个查询它所包含的不同语义主题

的个数不会超过 5 个, 而本文最终是为每个查询选取 4 到 5 幅代表性图像, 因此我们先选出 5 个代表性图像簇, 这些图像簇的质量相对较高, 然后再从代表性图像簇中选取代表性图像. 如果保留的图像簇个数太多, 则会出现语义主题重复现象;

5) 选取代表性图像簇中心图像作为图像集的表征.

3　实验与分析

3.1　实验图像集描述

在本文的实验中, 实验图像集来自 Google 搜索引擎, 包括卡通、目标、动物等共 10 大类别, 大小为 30 G. 每个类别对应约 1 000 个查询, 每个查询对应一个子图像集, 该图像集包含约 200 幅图像, 来自 Google 搜索引擎返回的约前 200 个结果.

对于每个查询, 本文将选取若干代表性图像作为该查询的表征. 为了验证本文算法的有效性和有用性, 采用其中 20 个查询作为实验对象, 它们分别对应 10 个不同的类别.

3.2　实算法比较

由于本文产生代表性图像是基于聚类的方法, 为了验证本文算法的有效性, 我们采用同样基于聚类的现有方法作为比较, 其中的两种方法如下:

1) 基于 K-means 聚类的方法: 在文献 [11] 中, 作者采用网格颜色矩特征来表示图像的空间颜色分布, 并采用 Gabor 纹理来表示纹理特征. 基于这两种全局特征进行 K-means 聚类, 局部特征 (SIFT 描述符) 用于图像簇排序.

2) 基于 Greedy K-means 聚类的方法: 在文献 [10] 中, 作者提取了 SIFT 共生特征, 采用球面 K-means 聚类方法将图像分成若干簇, 并利用贪婪算法从图像簇中选取代表性图像.

3.3　系统环境与评价标准

实验在 Matlab R2008a 环境下进行, 引入两种评价标准: 1) 查准率 (Precision), 即基于图像聚类的结果, 从而评价聚类算法的准确率; 2) 语义主题的查全率 (Recall of semantic theme), 即计算对于某个查询, 算法实际返回的代表性图像所包含的语义主题的个数与源图像集中所包含的语义主题总个数的比值. 具体定义如下:

假设某个查询的聚类结果中包含 N 个图像簇, 其中 C_i 表示第 i 个图像簇. 如果在一个图像簇中, 包含某个语义主题的图像个数最多, 那么该主题就是这个图像簇的主导主题. 设 $n_{C_i}^k$ 表示 C_i 中主导主题 k 的个数, N_{C_i} 表示中的图像总数, m 表示由算法所选取的代表性图像所包含的语义主题的个数, 源图像集中所包含的不同语义主题的类别数为 M, 则

$$Precision = \frac{1}{N}\sum_{i=1}^{N}\frac{n_{C_i}^k}{N_{C_i}} \tag{2}$$

$$Recall of semantic theme = \frac{m}{M} \tag{3}$$

其中, $k = 1, 2, \cdots, n$.

代表性图像的评价真实数据由人工评估者定义: 如果选取的图像集合包含了与该查询相关的多种不同语义主题, 则认为该图像集合是具有代表性的; 反之不具有代表性.

3.4　实验结果与分析

实验 1. 不同的 K 值对聚类准确率的影响

为了研究不同的 K 对聚类效果的影响, 以选取合适的 K 值, 我们选取了 5 个图像集作为实验对象, 计算 K 取不同值时查准率的大小. 实验结果如下图所示:

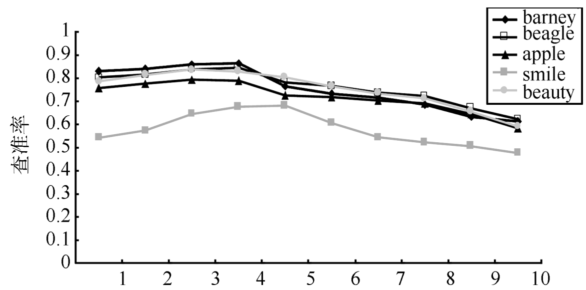


图 4　不同 K 值对查准率的影响

Fig. 4　Different K on the influence of precision

从图 4 中可以看出, 当 K 的取值大于 4 时, 大部分聚类结果的准确率开始下降, 即聚类效果不理想; 而当 K 值小于 4 时, 查准率呈上升的趋势, 因此 K 值取 4 时聚类效果最佳.

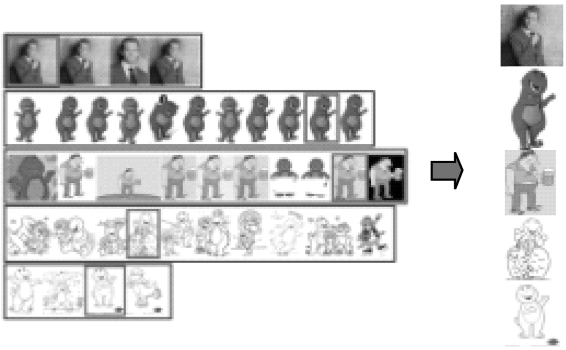
实验 2. 代表性图像选取

首先, 我们将展示利用本文方法对其中 3 个查询产生的实验结果, 如图 5 所示. 图 6 和图 7 是本文方法与现有方法的性能比较.

在图 5 中, 每个查询结果的每一行长矩形框内的图像集合都代表一个由本文算法选取的代表性图像簇. 而在每个代表性图像簇中用矩形框标出的单个图像就是该图像簇中的代表性图像. 代表性图像的集合如右侧图所示, 它包含了与该查询相关的不同语义主题的图像集合.

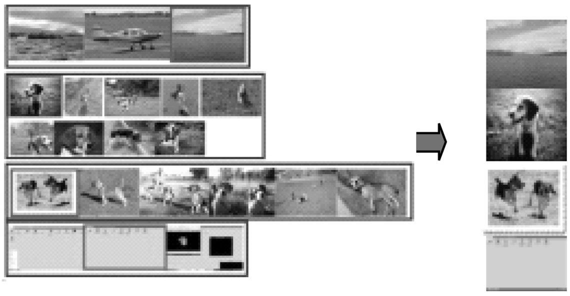
查询 "barney" 的实验结果如图 5 (a) 所示. 源图像集中共包含 4 种不同语义主题: "barney stinson"、"barney cartoon"、"barney dolls" 和 "barney gumble". 而从图中可以看出, 算法实际返回的 4 种不同语义主题的图像 (其中第 4 和第 5 图像簇

属于同一个类别 “barney cartoon”) 包含了源图像集中的所有语义主题, 即本文算法产生的代表性图像集合能够很好地描述源图像集.



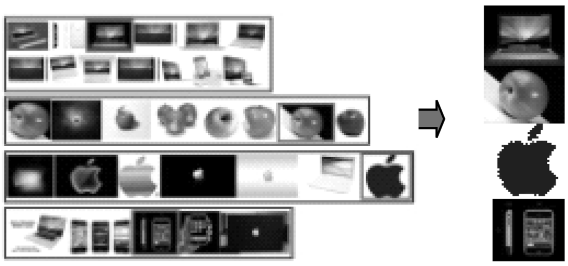
(a) 查询 “barney”

(a) Query “barney”



(b) 查询 “beagle”

(b) Query “beagle”



(c) 查询 “apple”

(c) Query “apple”

图 5 对三种查询产生的代表性图像结果

Fig. 5 Representative images selected for three queries

查询 “beagle” 的实验结果, 如图 5 (b) 所示. 结果包含了 3 种不同语义主题的图像, 分别是: “beagle channel”、“beagle dog” 和 “the interface of beagle development board”. 与源图像集相比, 缺少一种语义主题 “Beagle aircraft manufacturers”, 由于该语义图像背景较为复杂, 在提取特征时易受到背景的干扰, 因此不易被检索出来. 从总体上看, 本文算法能够返回大多数代表性图像.

查询 “apple” 实验结果如图 5 (c) 所示. 图中可以看出, 算法实际返回的语义主题有 4 种: “apple computer”、“apple fruit”、“apple logo” 和 “iphone”. 源图像集中还包含 “ipod” 和 “Steve Jobs”. 由于源图像集中包含 “ipod” 的图像较少 (共 4 幅) 且轮廓与 “iphone” 相类似, 在聚类过程中易被错分到 “iphone” 所属的图像簇中. 而人脸用底层的方法比较难以识别, 因此 “Steve Jobs” 也不易被检索出来. 同样地, 本文在总体上能够返回大多数代表性图像.

图 6 和图 7 中, 横轴代表选取的 20 个查询, 竖轴分别代表相应的查准率和语义主题的查全率. 最右边一列所示的是所有查询的评价标准的平均值. 其中 K-means 代表基于 K-means 聚类的方法, Greedy K-means 代表基于 Greedy K-means 聚类的方法, KMNC-AP 代表本文的方法. 从图中可以看出, 本文方法的查准率和语义主题的查全率均优于其他两种方法.

图 6 中, K-means 方法查准率的平均值是 0.731, 本文方法查准率的平均值相对于 K-means 方法提高了 6.2 %. 从单个查询看, 绝大部分聚类的性能较优. 在 Greedy K-means 方法中, 只利用了 SIFT 特征对图像进行聚类. 而单用一种特征不能够准确得描述图像内容, 因此, 其查准率低于本文方法的查准率.

在图 7 中, 对于查询 “barney”, 本文方法的语义主题查全率为 1, 即返回了源图像集中的所有 4 种语义主题, 而 K-means 方法为 0.5, 只返回了其中 2 种. 类似地, Greedy K-means 方法为 0.75, 只返回了其中 3 种, 所以其他两种方法均没有返回所有的语义主题. 从图中还可以看出, Greedy K-means 方法的性能要优于 K-means 方法. 从总体上看, 本文方法的查全率的平均值大于其他两种方法. 从单个查询看, 基本上可以选出包含多种语义主题的代表性图像. 在 Greedy K-means 方法中, 定义的代表性图像的质量函数容易受贪婪算法迭代过程中参数 α 和 β 的影响, 而 K-means 方法由于聚类的结果不够理想直接影响代表性图像的选取.

4 结论

本文针对传统图像检索系统返回的结果缺乏语义主题多样性问题, 提出了一种基于互近邻一致性和近邻传播的代表性图像选取算法. 首先利用互近邻一致性对图像相似度进行调整, 进而利用 AP 算法在特征空间中进行聚类, 最后, 根据文中提出的两种标准对图像簇进行排名选取代表性簇, 并从中选择中心图像作为代表性图像. 实验表明, 本文的方法在查准率和查全率方面的性能超过 K-means 方法和 Greedy K-means 方法. 本文算法主要有两个步

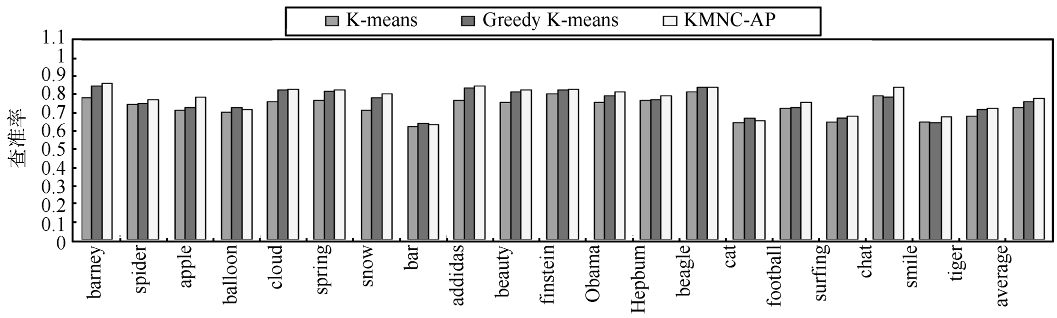


图 6　三种方法的查准率比较

Fig. 6　Comparison of the precision for three methods

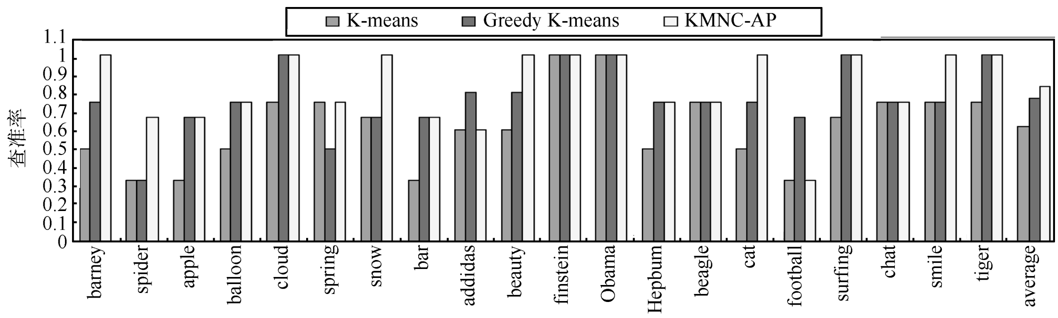


图 7　三种方法的语义主题查全率比较

Fig. 7　Comparison of the recall of semantic theme for three methods

骤比较耗时, 一个是提取图像的 SIFT 特征点以及词袋的生成, 另一个是 AP 聚类过程. 由于本文算法主要是用于离线计算, 所以对于时间的要求不高; 本文算法所需存储空间较大之处在于图像提取的 3 种特征以及聚类结果和代表性图像结果, 而算法程序占用空间较小. 从总体上看, 算法复杂度不高, 有可行性. 该方案可应用于搜索引擎结果和个人图片库的推荐浏览和查询.

另一方面, 本文的方法还存在一些不足: 1) 对于多种特征的融合采用的是加权累加. 然而对于不同的图像集, 各个特征的权重取值并不相同, 因此还需要设计一种自适应的权重分配方法, 能够有效地根据图像集自身的特性来分配各特征的权重, 从而更加有效地表示相似度量; 2) 本文完全采用无监督的方法产生代表性图像, 结果有时不尽人意. 如何转向有监督的方法, 提前对样本进行学习, 提高准确度, 是值得进一步研究的课题. 下一步将针对以上问题进行研究, 实现一种更优的大规模图像集中的代表性图像选取方法.

References

1 Wang M, Yang K Y, Hua X S, Zhang H J. Towards a relevant and diverse search of social images. *IEEE Transactions on Multimedia*, 2010, **12**(8): 829−842

2 Zha Z J, Yang L J, Mei T, Wang M, Wang Z F, Chua T S, Hua X S. Visual query suggestion: towards capturing user intent in internet image search. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2010, **6**(3): 1−19

3 Tang X O, Liu K, Cui J Y, Wen F, Wang X G. IntentSearch: capturing user intention for one-click internet image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(7): 1342−1353

4 Fan J P, Kem D A, Gao Y L, Luo H Z, Li Z M. JustClick: personalized image recommendation via exploratory search from large-scale flickr images. *IEEE Transactions on Circuits and Systems for Video Technology*, 2009, **19**(2): 273−288

5 Gao Y, Tang J H, Hong R, Dai Q H, Chua T S, Jain R. W2Go: A travel guidance system by automatic landmark ranking. In: Proceedings of the international conference on Multimedia. New York, USA: ACM, 2010. 123−132

6 Hong R, Tang J H, Tan H K, Ngo C W, Yan S C, Chua T S. Beyond search: event driven summarization for web videos. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2011, **7**(4): 1−21

7 Yang L J, Hanjalic A. Supervised reranking for web image search. In: Proceedings of the international conference on Multimedia. New York, USA: ACM, 2010. 183−192

8 Hama H, Zin T T, Tin P. A hybrid ranking of link and popularity for novel search engine. *International Journal of Innovative Computing, Information and Control*, 2009, **5**(11): 4041−4049

9 Jaffe A, Naaman M, Tassa T, Davis M. Generating summaries for large collections of geo-referenced photographs. In: Proceedings of the 15th international conference on World Wide Web. New York, USA: ACM, 2006, 853−854

10 Simon I, Snavely N, Seitz S M. Scene summarization for online image collections. In: Proceedings of IEEE 11th International Conference on Computer Vision. Rio de Janeiro: IEEE, 2007, 1−8

11 Kennedy L S, Naaman M. Generating diverse and representative image search results for landmarks. In: Proceeding of the 17th international conference on World Wide Web. New York, USA: ACM, 2008, 297−306

12 Raguram R, Lazebnik S. Computing iconic summaries of general visual concepts. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Anchorage, AK: IEEE, 2008. 1−8

13 Berg T L, Berg A C. Finding iconic images. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Miami, FL: IEEE, 2009, 1−8

14 Berg T L, Forsyth D A. Automatic Ranking of Iconic Images, Technical report, EECS Department, University of California, Berkeley, 2007

15 Chen Zeng, Hou Jin, Zhang Deng-Sheng, Zhang Hua-Zhong. Image re-ranking based on extraction of semantic regions. *Acta Automatica Sinica*, 2012, **37**(11): 1356−1359
(陈曾, 侯进, 张登胜, 张华忠. 基于语义区域提取的图像重排. 自动化学报, 2012, **37**(11): 1356−1359)

16 Frey B J, Dueck D. Clustering by passing messages between data points. *Science*, 2007, **315**(5814): 972−976

17 Ding C, He X F. K-nearest-neighbor consistency in data clustering: incorporating local information into global optimization. In: Proceedings of the 2004 ACM symposium on Applied computing. New York, USA: ACM, 2004, 584−589

18 Harel J, Koch C, Perona P. Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 2007, **19**: 545−552

19 Zhang Lin-Bo, Wang Chun-Heng, Xiao Bai-Hua, Shao Yun-Xue. Image representation using bag of phrases. *Acta Automatica Sinica*, 2012, **38**(1): 46−54
(张琳波, 王春恒, 肖柏华, 邵允学. 基于 Bag-of-phrases 的图像表示方法. 自动化学报, 2012, **38**(1): 46−54)

20 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91−110

21 Lin Hai-Feng, Ma Yu-Feng, Song Tao. Research on object tracking algorithm based on SIFT. *Acta Automatica Sinica*, 2010, **36**(8): 1204−1208
(蔺海峰, 马宇峰, 宋涛. 基于 SIFT 特征目标跟踪算法研究. 自动化学报, 2010, **36**(8): 1204−1208)

齐美彬 合肥工业大学计算机与信息学院教授. 主要研究方向为视频编码, 运动目标检测与跟踪和 DSP 技术.
E-mail: qimeibin@163.com
(**QI Mei-Bin** Professor at the School of Computer and Information, Hefei University of Technology. His research interest covers video coding, moving target detection and tracking and DSP technology.)

朱俊俊 合肥工业大学计算机与信息学院硕士研究生. 主要研究方向为多媒体检索, DSP 技术及应用. 本文通信作者.
E-mail: zhujunjun090523@126.com
(**ZHU Jun-Jun** Master student at the School of Computer and Information, Hefei University of Technology. Her research interest covers multimedia retrieval, DSP technology and applications. Corresponding author of this paper.)

纪 平 合肥学院电子信息与电子工程系讲师. 主要研究方向为信号与信息处理. E-mail: jiping@hfuu.edu.cn
(**JI Ping** Lecturer at the Department of Electronics & Electrical Engineering. Her research interest covers signal and information processing.)

蒋建国 合肥工业大学计算机与信息学院教授. 主要研究方向为数字图像分析与处理, 分布式智能系统和 DSP 技术及应用. E-mail: jgjiang@hfut.edu.cn
(**JIANG Jian-Guo** Professor at the School of Computer and Information, Hefei University of Technology. His research interest covers digital image analysis and processing, distributed intelligent systems and DSP technology and applications.)