基于兴趣目标的图像检索

张 峰,钟宝江
(苏州大学计算机科学与技术学院,江苏苏州 215000)

摘 要: 当前图像检索算法通常针对整体图像提取特征以完成检索任务. 然而,在很多情况下用户只会关注图像的一部分,即他们的兴趣目标. 此时,从整体图像提取的特征一部分是有效的,另一部分则是无效的且会对检索过程带来消极影响. 为此,本文提出基于兴趣目标的图像检索方案,并借助于现有的显著性检测、图像分割、特征提取等技术实现一款有效的图像检索算法. 首先采用 HS(Hierarchical Saliency,分层显著性)检测算法分析用户的兴趣目标并应用 SC(Saliency-based Image Cut,基于显著性的图像分割)算法将其分割,然后针对兴趣目标提取 HSV(Hue、Saturation、Value,色调、饱和度、明度)颜色特征、SIFT(Scale Invariant Feature Transform,尺度不变特征变换)局部特征和 CNN(Convolutional Neural Network,卷积神经网络)语义特征,最后计算其与数据库图像的相似度并根据相似度排序返回检索结果. 仿真实验结果表明,本文算法在解决"这是什么东西"这类图像检索任务时明显优于现有的图像检索算法.

关键词: 图像检索; 兴趣目标; 显著性检测; 图像分割; 卷积神经网络; 尺度不变特征变换; 颜色特征
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2018)08-1915-09
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2018.08.016

Image Retrieval Based on Interested Objects

ZHANG Feng, ZHONG Bao-jiang
(*Department of Computer Science and Technology, Soochow University, Suzhou, Jiangsu* 215000, *China*)

Abstract: The current image retrieval algorithms usually extract features from the whole input image to conduct retrieval tasks. However, in many cases users focus on only a part of the image, i.e. object-of-interest. As a result, the features extracted from the image are partially effective. In other words, some of the features are ineffective and might have a negative impact on the retrieval process. To overcome this difficulty, an image retrieval scheme based on object-of-interest is proposed. By incorporating this retrieval scheme with the existing techniques in saliency detection, image segmentation, and feature extraction, an effective image retrieval algorithm is coded. First, the hierarchical saliency (HS) detection algorithm is adopted to analyze the user's object-of-interest, and the saliency-based image cut (SC) algorithm is employed to segment it from the input image. Then, we extract the hue, saturation, value (HSV) color features, the scale invariant feature transform (SIFT) local features and the convolutional neural network (CNN) semantic features of the object-of-interest. Finally, the similarity of object-of-interest between a query image and every database image is computed and the retrieval result is sorted accordingly. Simulation experimental results show that, when being used to cope with a retrieval task like "what is this", the proposed algorithm is significantly better than the current image retrieval algorithms.

Key words: image retrieval; object-of-interest; saliency detection; image segmentation; convolutional neural network (CNN); scale invariant feature transform (SIFT); color feature

1 引言

在当前图像检索模型[1~5]与相应构建的图像搜索引擎中,通常针对整体图像来提取图像的底层特征. 从技术角度来说,这一做法是自然的,而从应用角度来说则可能无法满足人们的多种实际需求. 例如我们以"这是什么东西"为关键词在网络上取得一幅图像,见图 1(*a*) 所示. 显然,上传者本意是查询图像中蓝色粒状物体是什么东西. 然后我们将这幅图像分别上传至百度图像搜索引擎和 Google 图像搜索引擎进行查询识别. 这两款搜索引擎返回的前 10 幅检索结果分别如图 1(*b*) 和图 1(*c*) 所示. 可以看出,这些返回的图像在背景上(手掌)与查询图像类似并且也具有相同的语义(手掌托小物体),但均无助于用户解决其实际问题(识别用户所关注的兴趣目标,即蓝色粒状物体). 可见,如果用户仅仅对图像中的特定目标物体感兴趣,此时从整

收稿日期:2017-02-03;修回日期:2018-01-18;责任编辑:蓝红杰
基金项目:国家自然科学基金(No. 61572341);苏州大学"东吴学者计划"项目

体图像提取的特征一部分是有效的,而另一部分则是无效的且会对查询结果带来消极影响. 为此,本文提出基于兴趣目标的图像检索模型,用以解决图像搜索引擎进行“这是什么东西”之类的查询任务.

显著性检测是图像分析与理解领域另一个重要分支,其主要任务是模拟人眼视觉注意选择机制,检测出图像中密度、颜色、形状等与周围区域有显著差异的区域[6~10]. 本文中,当进行“这是什么东西”这类查询任务时,我们以图像显著性检测来理解用户的兴趣目标. 这是因为用户一般会有意识地将其所关注的目标用醒目的方式来呈现并希望计算机能够理解其意图. 这是人类之间交流的一种正常方式,比如图 1(*a*) 中,用户的兴趣目标(蓝色粒状物体)与背景(手掌)差异很大. 很难理解人们会将需识别的兴趣目标混杂在类似的物体中进行查询.

(*a*) 查询图像

(*b*) 百度搜索引擎的结果

(*c*) Goolge 搜索引擎的结果

图1 当前图像搜索引擎的检索结果

此前,研究者已经尝试将视觉注意机制融入到图像检索框架中. Fu 等人[11]提出了基于注意力驱动的图像检索系统,该方法将显著物体从背景中分离出来,并赋予较高的注意值;检索时,只比较注意值较高的目标物体. Liu 等人[12]提出了一种利用显著性结构直方图描述图像的方法. 该方法融入视觉注意内核和神经元的方向选择性机制,以此来提高检索系统的准确性. 然而,在 Fu 等人[11]提出其算法时,人们对显著性检测问题的理解还不够完善,相关技术效率较低,并且该算法所提取的图像特征仅包括颜色和纹理,在描述目标时区分力不够,从而导致图像检索效率低下. Liu[12]等人的算法介于基于整体图像的检索和基于兴趣目标的检索之间,可以解决目标类似(权重较高)同时背景类似(权重较低)的图像检索任务. 该算法从功能上来说与现有的图像检索引擎的表现类似,但存在的问题是如果用户所感兴趣的是图像中的特定目标,此时背景特征将降低检索效率.

本文将结合显著性检测与图像分割领域最新研究成果,实现一种基于用户兴趣目标的图像检索算法. 与 Fu 等人[11]提出的算法相比,本文算法的检索性能显著提升. 与基于整体图像的检索模型及 Liu 等人[12]提出的检索模型相比,本文算法解决了用户不同的检索任务. 新算法的主要思想是:依据 HS 显著性检测算法[10]分析用户的兴趣目标,结合 SC 算法[9]分割出兴趣目标;然后对用户的兴趣目标提取 HSV 颜色特征、SIFT 局部特征和 CNN 语义特征;最后将其与数据库图像进行特征相似度匹配,并根据相似度排序得到基于兴趣目标的检索结果. 以上仅在兴趣目标区域提取特征的做法,可以有效抑制背景对检索结果的影响,提高检索的精度和召回率. 流程如图 2 所示.

2 分层显著性模型

本节结合我们所考虑的问题来分析和测试常用显著性检测算法的性能,并确定采用分层显著性(HS)算法[10]来理解和获取用户的兴趣目标. HS 算法首先提取输入图像不同尺度的图像层,然后计算每张图像层的显著性线索,最后利用图模型将每层的显著性线索融合成一张显著图. 图 3 示意了 HS 算法的主要步骤,其中图 3(*a*)为输入图像.

2.1 提取图像层

图像层是对输入图像在不同细节程度上的描述,不同层对输入图像的表示和结构复杂度的表现不一样. 图像层的层数一般设定为 3 层. 在第 1 层,图像的细节尽可能被保留,在第 2 层,图像的细节消失,显现出图像的结构,在第 3 层,细节消失,只剩下大尺度的图像结构.

为了产生如上的三幅图像层,首先使用分水岭分割算法[13]生成一幅初始化的过分割图,并计算每个分割区域的尺度值. 然后使用一个迭代程序合并邻近的分割区域. 如果区域尺度值小于 3,这个区域将被合并到最近邻的区域,随之更新合并后区域的尺度值和颜色. 当所有的区域都处理完后,将产生第 1 层区域图. 第 2 层区域图和第 3 层区域图的产生方式与第 1 层类似,只是尺度阈值更大(如图 3(*b*) 所示).

2.2 单层显著性线索

用于单层显著性的主要线索包括局部对比度和启发式位置.

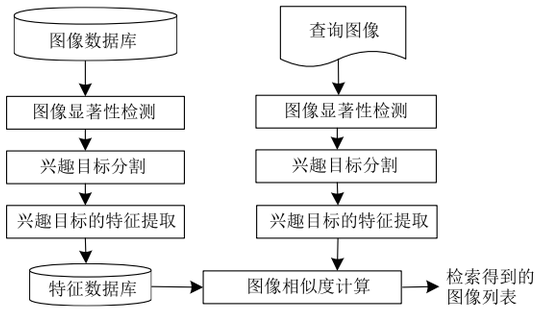
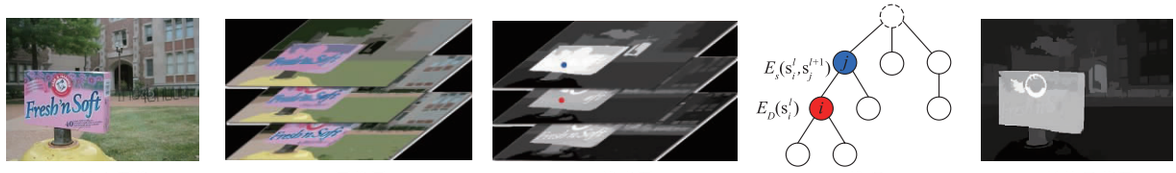


图2　基于兴趣目标的图像检索流程



(*a*) 输入图像　(*b*) 图像层　(*c*) 线索图　(*d*) 推理　(*e*) 最终显著图

图3　HS 算法步骤

(1)局部对比度　与周围颜色对比度较大的图像区域一般更会吸引人们的关注. 考虑图像的两个区域 R_i 和 R_j,其颜色分别记为 c_i 和 c_j. 区域的像素数目为 $w(R_j)$. 记:

$$\varphi(i,j)=\exp\{-D(R_i,R_j)/\sigma^2\}$$

为区域 R_j 在空间上对区域 R_i 的显著性影响程度,其 $D(R_i,R_j)$ 表示区域 R_i 中心和区域 R_j 中心的欧氏距离的平方,参数 σ^2 控制周围区域影响范围. 区域 R_i 局部对比度显著性线索定义如下:

$$C_i=\sum_{j=1}^{n} w(R_j)\varphi(i,j)\parallel c_i-c_j\parallel_2 \tag{1}$$

其中 n 为图像中区域的总数.

(2)启发式位置　心理学研究表明人们注意力倾向于图像中间区域[14],因此靠近图像中心的区域显著性更高. 记 x_c 为图像中心,$\{x_0,x_1,\cdots\}$ 为区域 R_i 中像素坐标的集合. 启发式位置模型如下:

$$H_i=\frac{1}{w(R_i)}\sum_{x_i\in R_i}\exp\{-\lambda\parallel x_i-x_c\parallel^2\} \tag{2}$$

为了更好地得到图像的显著性,需要融合以上形式线索,形式如下:

$$\bar{s}_i=C_i\cdot H_i \tag{3}$$

其中 λ 控制位置线索与局部对比度线索的权重. λ 越大,位置线索权重越小,一般 λ 设置为 9. 对每一层计算完线索融合值 $\bar{s}_i$ 后,可以得到初始的显著图(如图 3(*c*)所示).

2.3　分层推理

借助树结构的图模型进行分层推理,可以实现对所有线索图的融合(如图 3(*d*) 所示). 在第 $k(k=1,2,3)$ 层,对区域 i 对应的节点定义显著性变量 $s_i^{(k)}$,集合 S 包含所有的显著性变量. 为了分层推理,需要求解以下能量函数的最小化问题:

$$E(S)=\sum_k\sum_i E_D(s_i^{(k)})+\sum_k\sum_i E_s(s_i^{(k)},s_j^{(k+1)}) \tag{4}$$

其中第二项要求 $R_i^{(k)}\subseteq R_j^{(k+1)}$. 该函数包含两部分,分别为数据项和层次项. 数据项 $E_D(s_i^{(k)})$ 表示各层区域显著性置信度,定义如下:

$$E_D(s_i^{(k)})=\beta^{(k)}\parallel s_i^{(k)}-\bar{s}_i^{(k)}\parallel_2^2 \tag{5}$$

其中 $\beta^{(k)}$ 控制层置信度,并且 $\bar{s}_i^{(k)}$ 是由式(3)计算得到的初始化的显著性值.

层次项 $E_s(s_i^{(k)},s_j^{(k+1)})$ 控制不同层对应区域的一致性. E_s 定义如下:

$$E_s(s_i^{(k)},s_j^{(k+1)})=\lambda^{(k)}\parallel s_i^{(k)}-s_j^{(k+1)}\parallel_2^2 \tag{6}$$

其中 $\lambda^{(k)}$ 控制层与层之间的一致性强度. 层次项使得不同层对应区域的显著性分配更相似,能够有效地纠正初始显著性错误.

式(4)中的能量函数是一个简单的分层图模型,采用置信传播[15]的方法可以实现最小化. 当能量函数达到全局最小时,便可得到最终的显著图(如图 3(*e*) 所示).

2.4　实验分析

验证一个显著性检测算法的性能最简单的方法是设置一个阈值 $T_f\in[0,255]$ 对算法产生的显著图进行二值化,从而得到兴趣目标的二值分割. 为了全面地比较各种显著性检测算法凸显兴趣目标的好坏,阈值 T_f 从 0 到 255 动态地变化. 我们根据二值化显著图与手工标注的目标显著性区域进行比较来评估,评估准则采用精度-召回率(Precision-Recall,PR)曲线.

图4 展示了本文所采用HS算法[10]结合SC算法[9]提取兴趣目标的结果,图4(a)为输入图像,图4(b)为HS算法检测得到的显著图,图4(c)是基于显著图分割出兴趣目标的结果,图4(d)是人工标注的兴趣目标区域.可以看出,最终的显著性区域均可以有效指向兴趣目标.图5给出了HS算法与其他4种显著性检测算法(IT[6],SR[7],HFT[8],RC[9])在SIVAL数据库[16]上的性能表现.可以看出,HS算法能够取得最佳的效果.

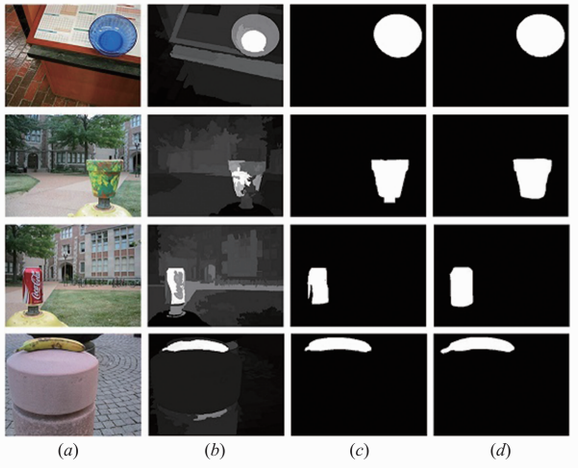


图4 结合HS算法和SC算法提取兴趣目标的结果 (a)输入图像; (b)显著图; (c)兴趣目标分割结果; (d)人工标注的兴趣目标

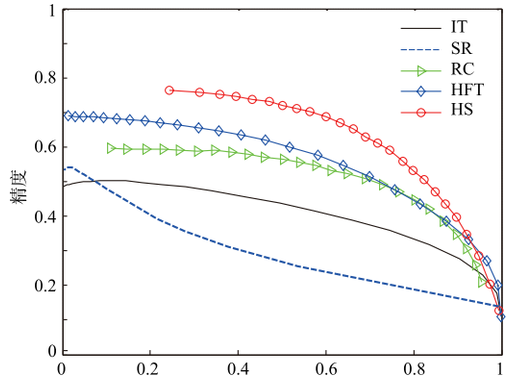


图5 不同显著性检测算法的结果比较,其中HS为本文所采用的算法

3 兴趣目标分割

本节测试和分析常用的图像分割算法,并确定采用SC算法[9]来分割用户的兴趣目标.在图像分割领域,SC算法是对GrabCut算法[17]的一种改进.GrabCut算法需要用户在图像中框选出所要分割的目标,而SC算法则利用显著性检测来理解用户期待的目标区域,从而不需要人工参与就能自动选择目标区域.可以看出,SC算法与本文算法有着类似的做法,即均使用了显著性检测来理解和获取用户的目标.不同的是,SC算法使用了RC显著性检测算法[9],而本文算法使用了效率更高的HS显著性检测算法[10].SC算法实现步骤如3.1节和3.2节所示.

3.1 兴趣区域初始化

SC算法首先对图像进行显著性检测,然后利用显著图来生成一个不完全的三值图(0表示背景像素,128表示未知像素,255表示目标像素).显著性值低于阈值的像素被认为背景像素,其余像素被认为可能是目标像素也有可能是背景像素,对应于三值图中的未知像素.此时三值图中值为255的像素个数为0,之后值为128的像素可被赋为255,因此此处三值图为不完全的三值图.初始三值图中的背景像素用来训练背景颜色模型,未知像素用来训练前景颜色模型[17].在生成不完全的三值图时,此算法将置信度非常高的非显著性区域中的像素作为背景像素.

3.2 兴趣区域分割

初始化步骤完成后,SC算法迭代地调用了GrabCut算法来改进兴趣区域的分割结果(一般最多迭代4次).每次迭代后,分别对分割结果使用膨胀和腐蚀操作来得到新的三值图以进行下一次迭代.膨胀后仍然落在外面的区域像素设成背景像素,腐蚀后仍然落在区域内的像素设为前景像素,其余像素为三值图中的未知像素.此时,背景像素用来训练背景颜色模型,前景像素用来训练前景颜色模型.SC算法流程如图6所示.

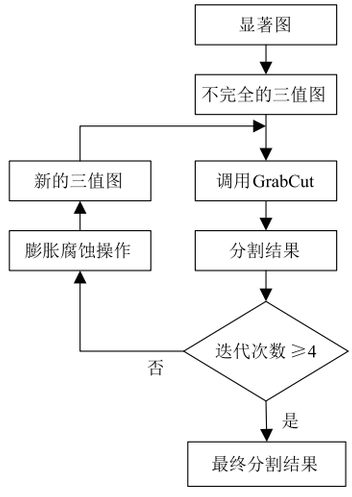


图6 SC算法流程图

3.3 实验分析

基于显著图分割出兴趣目标的常用方法是设定一个经验阈值 $T_f \in [0,255]$ 对显著图进行二值化.该方法称为固定阈值分割法(Fixed Threshold Cut,FTC).另一种常用方法是自动阈值分割法(Automatic Threshold Cut,ATC)[18].图7给出了SC算法与以上两种阈值分割算法的比较结果.分割效果的评估采用精度,召回

率,F-度量这三项指标. 从图 7 的 F-度量这项综合指标可以看出,SC 算法效果优于 ATC 算法和 FTC 算法,并且从实验结果观察来看,SC 算法获得的兴趣目标区域更加集中,不像阈值分割那么分散.

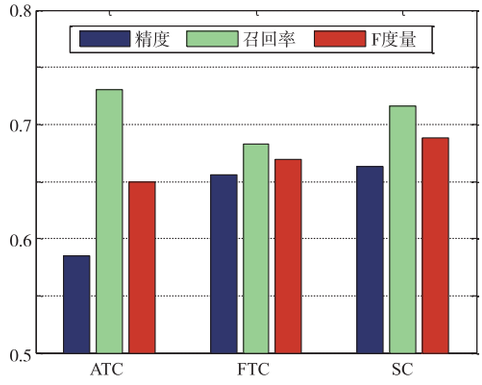
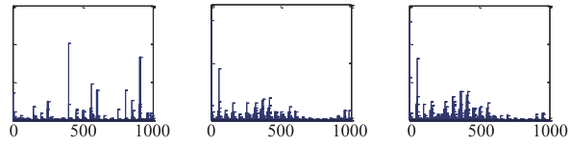
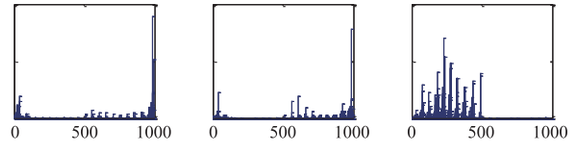


图7　不同的显著图分割方法的比较

(*a*) 测试图像



(*b*) 整体图像的颜色直方图



(*c*) 兴趣目标的颜色直方图

图8　颜色特征的重要性

4　兴趣目标的特征提取

在应用 HS 算法和 SC 算法理解和获取用户的兴趣目标后,为了对其进行描述,我们分析和测试了若干不同图像特征,最终确定使用 HSV 颜色特征、SIFT 局部特征和 CNN 语义特征相结合的方式从多个不同的角度去描述兴趣目标. 下面首先给出以上三种特征的提取细节,然后融合这些特征进行图像相似度计算.

4.1　兴趣目标的 HSV 颜色特征

由于 RGB 颜色空间与人眼的感知差异较大,因此本文采用更符合人眼感知特性的 HSV 颜色空间[19]. 首先根据兴趣目标分割结果,保留兴趣目标区域的像素. 接着将兴趣目标中所有像素的(r,g,b)值转换为(h,s,v)值,并将 HSV 颜色空间量化成 $20\times10\times5=1000$ 种颜色[2]. 最后,用归一化的 1000 维 HSV 颜色直方图描述兴趣目标的颜色特征. 图 8 示例了以上颜色特征对检索兴趣目标的意义. 图 8(a)给出了三幅测试图像,其中,前两幅图像包含相同的兴趣目标,但背景不同;后两幅图像背景相同,但包含不同的兴趣目标. 实验结果表明,若比较整体图像颜色直方图,后两幅图像较相似,见图 8(b);若比较兴趣目标的颜色直方图,则前两幅图像较相似,见图 8(c). 显然,应用 HSV 颜色特征能够有效地描述与识别兴趣目标.

4.2　兴趣目标的 SIFT 特征

鉴于 SIFT 特征的一系列优良特性,我们提取出兴趣目标的 SIFT 局部特征. 首先采用 DoG[20] 检测器检测出图像中稳定的关键点,并且根据兴趣目标分割结果保留兴趣目标区域的关键点,然后用 128 维向量描述兴趣目标区域内每个关键点周围 16×16 区域的信息. 本文在独立的数据集上(为保证视觉词典的通用性)训练得到 $20k$ 的视觉词典,随之将每个 SIFT 特征通过最近邻算法[21] 量化成视觉单词. 最终建立一个标准的倒排索引,并利用投票机制进行检索. 图 9 示例了 SIFT 特征对于兴趣目标匹配的重要性,比较图 9(a) 和 9(c) 可以看出,目标相同背景不同的图像进行匹配时,基于兴趣目标的匹配,能够有效去除由于背景干扰而产生的错配对;从图 9(b) 和 9(d) 比较中可以看出,背景相同目标不同的两幅图像,虽然在背景区域能够产生大量匹配对,但由于我们关注的是兴趣目标部分,所以背景区域的匹配对应该给予剔除.

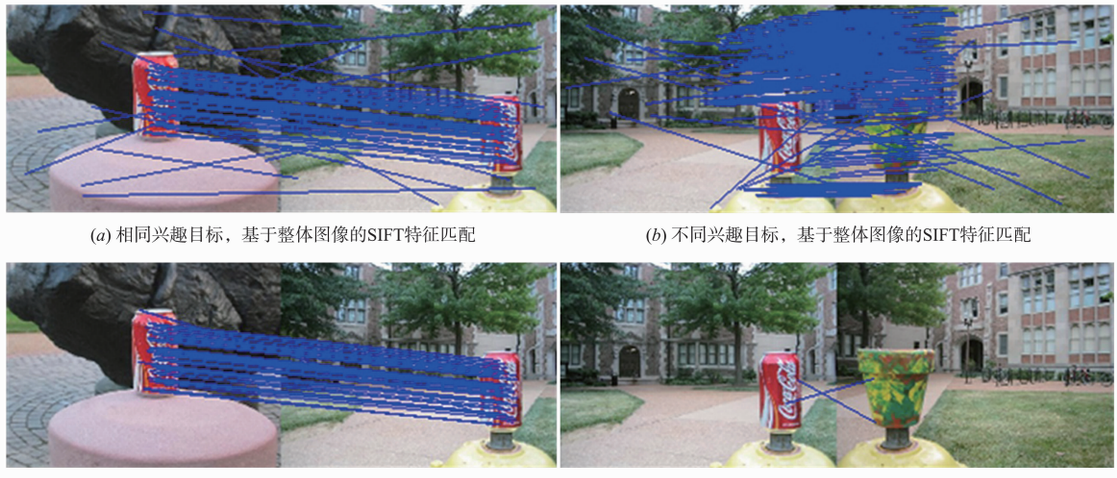
4.3　兴趣目标的 CNN 特征

卷积神经网络(Convolutional Neural Network, CNN)[22] 是一种多层神经网络模型. 在底层,提取的特征较原始,层次越高,提取的特征越抽象,在高层已经是一种语义组合. 这种网络结构提取的特征对平移变换、旋转变换、仿射变换等具有高度不变性. 为了提取兴趣目标的语义特征,本文根据兴趣目标的分割结果,用一个矩形框包含兴趣目标并将其剪切出来. 鉴于 GoogLeNet 网络模型[23] 运行速度快,提取的特征维度低,区分力度强,我们利用已训练好的 GoogLeNet 模型提取兴趣目标的 1024 维的 CNN 特征向量,并对该特征向量进行归一化. 仿真实验表明,提取剪切后兴趣目标的特征比提取整体图像的特征更能够描述图像的目标部分.

4.4　基于兴趣目标的图像相似度计算

通过分析和测试,本文使用加权的特征距离计算查询图像 Q 和数据库中每一幅图像 I 之间的相似度,然后按照相似度由大到小的顺序返回图像检索结果.

记 $S_h(Q,I)$ 为两幅图像兴趣目标的 HSV 颜色直方图相似度,计算如下:



(*a*) 相同兴趣目标，基于整体图像的SIFT特征匹配　(*b*) 不同兴趣目标，基于整体图像的SIFT特征匹配

(*c*) 相同兴趣目标，基于兴趣目标的SIFT特征匹配　(*d*) 不同兴趣目标，基于兴趣目标的SIFT特征匹配

图9　SIFT特征的重要性

$$S_h(Q,I)=\sum_{k=1}^{N}|h_k(Q)h_k(I)| \tag{7}$$

其中，N 为直方图区间数.

记 $S_s(Q,I)$ 为两幅图像的兴趣目标区域 SIFT 匹配数的得分，匹配点的数目越多，该得分越高，计算如下：

$$S_s(Q,I)=\frac{\sum_{\vec{x}\in Q,\vec{y}\in I} f(\vec{x},\vec{y})\cdot idf^2}{\|Q\|_2\ \|I\|_2} \tag{8}$$

其中 $\vec{x},\vec{y}$ 表示图像 Q 和 I 中 SIFT 特征，$f(\vec{x},\vec{y})$ 表示两 SIFT 特征的匹配函数，匹配则为 1，不匹配则为 0，idf 表示倒排文档频率，$\|Q\|_2$ 表示词频的欧式范数.

记 $S_c(Q,I)$ 为两幅图像兴趣目标的 CNN 特征相似度，采用余弦距离度量，计算如下：

$$S_c(Q,I)=\frac{h(Q)\cdot h(I)}{\|h(Q)\|_2\ \|h(I)\|_2} \tag{9}$$

基于以上三种特征相似度得分以及乘法法则融合策略[24]，查询图像 Q 和数据库中的图像 I 之间的相似度定义为：

$$S(Q,I)=S_h(Q,I)^{w_h}\cdot S_s(Q,I)^{w_s}\cdot S_c(Q,I)^{w_c} \tag{10}$$

其中 w_h,w_s,w_c 为上述 3 个特征对应的权值，满足 $w_h+w_s+w_c=1$. 基于仿真实验结果，这些参数默认值取为 $w_h=0.28$，$w_s=0.02$，$w_c=0.7$.

5　实验结果及分析

以下首先介绍实验选择的数据库和测评指标，然后分别验证 HSV，SIFT，CNN 作为单一特征对检索兴趣目标的优越性以及多特征组合对检索的有效性，最后一部分验证本文算法的检索性能.

5.1　实验数据库选择

我们选择与问题背景相符的 SIVAL 图像数据库[16]来展示和评估本文算法的性能. 该数据库共由 1500 幅图像组成，分为 25 类，每类 60 幅图像. 同类图像均含有一个相同的目标，但其背景具有高度多样性，且目标的空间位置、尺度大小、光照等在不同的图像中也会发生很大的变化. 该数据库将目标相同的图像归为一类，因此检索时需要忽略图像背景而关注对目标的描述和识别. 图 10 展示了 SIVAL 数据库的部分样例.

图10　SIVAL 数据库的样例图像

在综合测评本文算法时，为了验证算法的稳定性，我们将 MIRFLICKR-25k 数据库[25]作为干扰类图像与 SIVAL 数据库合并，形成一个具有 26 类共 26500 幅图像数据库，称为“SIVAL-MIRFLICKR”数据库. 在这一融合的图像数据库上，我们以其中 SIVAL 数据库的各幅图像作为查询图像，通过统计查询准确率和召回率等指标来评估本文算法和多个对比算法的检索性能.

5.2　评估指标

评估图像检索性能主要有两个指标：精度(Precision)和召回率(Recall). F-度量(F-Measure)为这两个指标的调和平均数，是对精度和召回率综合性能的评估. 精度 P、召回率 R 以及 F-度量具体计算方法如下：

$$P = I_N / N$$
$$R = I_N / M$$
$$F = \frac{(1+\beta^2) \times P \times R}{\beta^2 \times P + R} \tag{11}$$

其中 I_N 为检索返回的同类图像数目,N 为检索返回的图像数目,M 为数据库中所包含的同类图像数目. 参数 β 权衡精度和召回率之间的重要性.

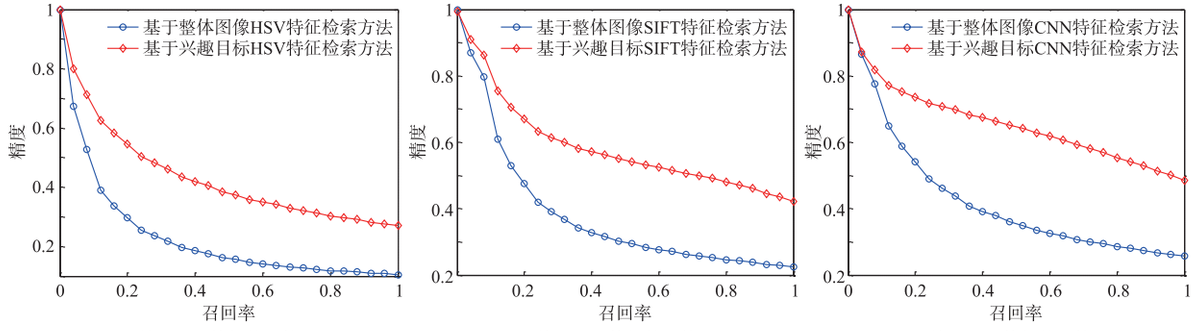
若用 X 轴表示精度,Y 轴表示召回率,可得到精度-召回率(Precision-Recall,PR)曲线. 平均精度(Mean Average Precision,mAP)一般用来度量 PR 曲线变化的差异,计算如下:

$$\text{mAP} = \frac{\sum_{q=1}^{Q} \sum_{k=1}^{n} p(k) \Delta r(k)}{Q} \tag{12}$$

其中 $p(k)$ 和 $r(k)$ 分别对应精度和召回率,Q 为查询样本的数目. 平均精度即为 PR 曲线与 X 轴 Y 轴所围成的面积.

5.3　基于兴趣目标与基于整体图像的检索效果比较

兴趣目标的每一种特征都可以单独用于匹配检索. 为了展现单一特征对最终检索结果的贡献度,我们首先提取整体图像和兴趣目标的 HSV、SIFT 和 CNN 特征. 对于不同特征,分别比较基于兴趣目标的检索算法和基于整体图像的检索算法. 图 11 给出了检索结果的 PR 曲线图. 可以看出,对于不同特征,基于兴趣目标的检索算法都能够大幅度提升检索效果. 相比于 SIFT 特征和 CNN 特征,HSV 特征对检索效果提升的幅度稍低. 这是因为颜色特征容易受到光照、噪声等因素影响. CNN 语义特征和 SIFT 局部特征都具有优秀的区分力,在去除图像背景区域的影响后,对兴趣目标的检索性能均得到了较大提升.



(*a*) 关于HSV特征的比较　(*b*) 关于SIFT特征的比较　(*c*) 关于CNN特征的比较

图11　基于兴趣目标与基于整体图像的检索效果比较

5.4　多特征组合与单一特征检索效果比较

单一的特征可能在某些样本上效果比较好,但在另外一些样本上效果会变差. 为此,一般通过多特征组合的方式来进行图像检索. 为了研究不同特征组合的贡献,本文分别提取图像中兴趣目标的 HSV、SIFT 和 CNN 特征,并且比较这三种特征以及它们组合特征的检索效果. 根据实验结果(见图 12)可以得出两个主要结论:(1)仅仅使用单一特征很难获得较好的检索效果. 在实验中,CNN 特征的检索效果最好,优于 SIFT 特征和 HSV 特征. (2)两个或者三个特征的组合,会明显提高检索效果. 三个特征相组合的检索效果达到最佳.

5.5　本文算法与现有检索算法的比较

依据 5.4 节的实验结果与分析,本文选取 HSV,SIFT,CNN 三种特征的联合形式来描述兴趣目标,然后通过相似度计算完成基于兴趣目标的图像检索. 为了可客观地测评新算法,我们遴选了 HE[1]、c-MI[2]、CDH[3]、MSD[4]、SSH[12]、NC[26]检索算法作参照.

图 13 展示了本文算法和基于整体图像的算法(以 NC 算法为代表)在 SIVAL 数据库上的检索样例,对检索结果的观察可以看出,在返回的前 24 张图像中,本文算法检索出的图像中的兴趣目标均与查询图像中的兴趣目标(花盆)相同,见图 13(*a*) 所示,而 NC 算法检索出的图像均在背景上(蓝色椅子)与查询图像相同,见图 13(*b*) 所示. 可见,基于整体图像的算法无助于解决用户的实际问题.

图 14 和表 1 分别给出了新算法与比较算法在 SIVAL 数据库和 SIVAL-MIRFLICKR 数据库上测试的 PR 曲线图和具体的数值结果. 可以看出,本文提出的算法效果明显高于其他算法. 分析这些比较算法可知,这些算法都是提取整体图像的特征,在进行检索时,检索效果严重受到了背景的干扰. 而本文算法结合显著性检测与图像分割技术来理解和获取用户的兴趣目标,并且仅针对兴趣目标进行特征描述. 因而本文算法对解决“这是什么东西”这类检索任务具有较大优势,弥补了现有的图像检索算法的不足.

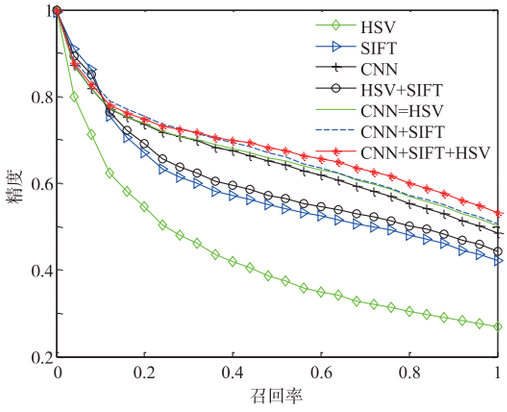
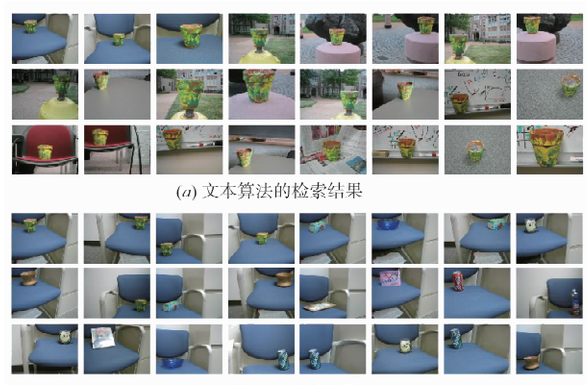


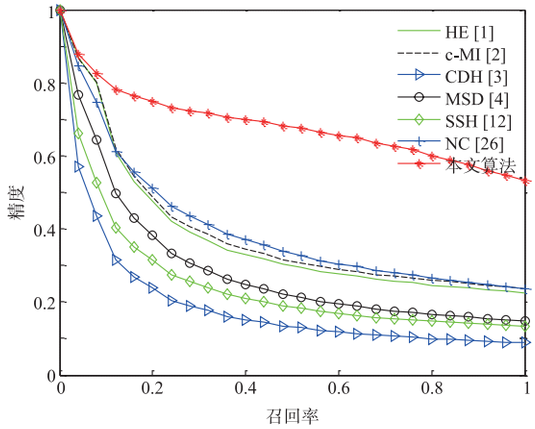
图12 不同特征相结合检索效果的比较



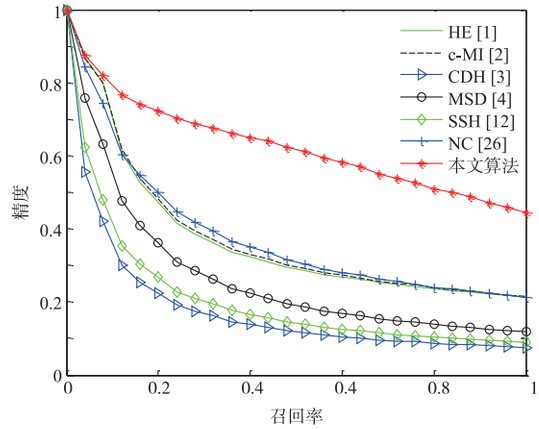
(*a*) 文本算法的检索结果

(*b*) NC算法的检索结果

图13 文本算法和NC算法的检索样例



(*a*) SIVAL数据库



(*b*) SIVAL-MIRFLICKR数据库

图14 本文算法与现有其他算法的比较

表1 本文算法与现有其他算法检索结果

数据库\算法	HE[1]	c-MI[2]	CDH[3]	MSD[4]	SSH[12]	NC[26]	本文算法
SIVAL,mAP	0.286	0.284	0.093	0.162	0.131	0.263	**0.56**
SIVAL-MIRFLICKR,mAP	0.234	0.223	0.055	0.097	0.067	0.204	**0.449**

6 结论

用户使用现有的图像搜索引擎查询一幅图像时,可能是想识别图像中所感兴趣的目标或者返回与兴趣目标相关的图像.然而,当前图像检索算法大多根据整体图像的特征来设计,很难满足这种需求.本文针对这一课题开展研究,提出了基于兴趣目标的图像检索方案.首先采用HS算法和SC算法来理解和获取用户的兴趣目标,然后选取兴趣目标的HSV颜色特征、SIFT局部特征和CNN语义特征进行基于兴趣目标的图像检索,最终实现了可实际用于解决"这是什么东西"这类查询任务的检索算法.实验结果表明,本文提出的算法与基于整体图像的检索算法相比,在解决兴趣目标识别的任务上具有更佳的性能.

参考文献

[1] Jegou H, Douze M, Schmid C. Hamming embedding and weak geometric consistency for large scale image search [A]. European Conference on Computer Vision[C]. Berlin, Germany: Springer, 2008. 304 – 317.

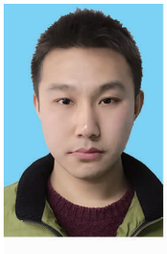
[2] Zheng L, Wang S, Liu Z, et al. Packing and padding coupled multi-index for accurate image retrieval [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. New York, USA: IEEE, 2014. 1947 – 1954.

[3] Liu G H, Yang J Y. Content-based image retrieval using color difference histogram [J]. Pattern Recognition, 2013, 46(1): 188 – 198.

[4] Liu G H, Li Z Y, Zhang L, et al. Image retrieval based on micro-structure descriptor [J]. Pattern Recognition, 2011,

44(9):2123 – 2133.
[5] 周燕,曾凡智. 基于二维压缩感知和分层特征的图像检索算法[J]. 电子学报,2016,44(2):453 – 460.
Zhou Yan, Zeng Fan-zhi. An image retrieval algorithm based on two-dimensional compressive sensing and hierarchical feature[J]. Acta Electronica Sinica, 2016, 44(2): 453 – 460. (in Chinese)
[6] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Trans on PAMI, 1998, 20(11): 1254 – 1259.
[7] Hou X, Zhang L. Saliency detection: A spectral residual approach[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. New York, USA: IEEE, 2007. 1 – 8.
[8] Li J, Levine, M. D, An X, et al. Visual saliency based on scale-space analysis in the frequency domain[J]. IEEE Trans on PAMI, 2013, 35(4): 996 – 1010.
[9] Cheng M M, Zhang G X, Mitra N. J. , et al. Global contrast based salient region detection[J]. IEEE Trans on PAMI, 2015, 37(3): 409 – 416.
[10] Yan Q, Xu L, Shi J, et al. Hierarchical saliency detection[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. New York, USA: IEEE, 2013. 1155 – 1162.
[11] Fu H, Chi Z, Feng D. Attention-driven image interpretation with application to image retrieval[J]. Pattern Recognition, 2006, 39(9): 1604 – 1621.
[12] Liu G H, Yang J Y, Li Z Y. Content-based image retrieval using computational visual attention model[J]. Pattern Recognition, 2015, 48(8): 2554 – 2566.
[13] Gonzalez R C, Woods R E. Digital Image Processing[M]. Nueva Jersey, 2008.
[14] Tatler B W. The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions[J]. Journal of Vision, 2007, 7(14): 1 – 17.
[15] Kschischang F R, Frey B J, Loeliger H A. Factor graphs and the sum-product algorithm[J]. IEEE Trans on Information Theory. 2001, 47(2): 498 – 519.
[16] http://www. cse. wustl. edu/ ~ sg/multi-inst-data/[OL]
[17] Rother C, Kolmogorov V, Blake A. "GrabCut": interactive foreground extraction using iterated graph cuts[J]. ACM Trans on Graphics, 2004, 23(3): 309 – 314.
[18] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Trans on Systems Man and Cybernetics, 1979, 9(1): 62 – 66.
[19] Paschos G. Perceptually uniform color spaces for color texture analysis: an empirical evaluation[J]. IEEE Trans on Image Processing, 2001, 10(6): 932 – 937.
[20] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
[21] Muja M, Lowe D G. Scalable nearest neighbor algorithms for high dimensional data[J]. IEEE Trans on PAMI, 2014, 36(11): 2227 – 2240.
[22] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 2012.
[23] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. New York, USA: IEEE, 2015. 1 – 9.
[24] Alkoot F M, Kittler J. Experimental evaluation of expert fusion strategies[J]. Pattern Recognition Letters, 1999, 20(11): 1361 – 1369.
[25] Huiskes M J, Lew M S. The MIR flickr retrieval evaluation[A]. ACM International Conference on Multimedia Information Retrieval[C]. ACM, 2008. 39 – 43.
[26] Babenko A, Slesarev A, Chigorin A, et al. Neural codes for image retrieval[A]. European Conference on Computer Vision[C]. Cham, Germany: Springer, 2014. 584 – 599.

作者简介



张　峰　男,1990 年生于江苏扬州. 苏州大学计算机科学与技术学院硕士研究生. 研究方向为图像处理、图像检索.
E-mail:276197179@ qq. com

钟宝江(通信作者)　男,1972 年生于江苏盐城. 苏州大学计算机科学与技术学院教授. 研究方向为计算机视觉、图像处理、数值线性代数等.
E-mail:bjzhong@ suda. edu. cn