

一种面向草图检索的边缘特征提取算法

王若梅^{1,2)}, 姜俊锬^{1,2)}, 林淑金^{2,3)*}, 汪 飞^{1,2)}

¹⁾ (中山大学数据科学与计算机学院 广州 511400)

²⁾ (中山大学国家数字家庭工程技术研究中心 广州 511400)

³⁾ (中山大学传播与设计学院 广州 511400)

(linshujin@mail.sysu.edu.cn)

摘要: 面对图像数据的激增,如何高效地检索到图像信息仍然是研究人员的热点研究问题,目前已形成了以图搜图为主流的检索方式.随着触摸屏人机交互技术的兴起,基于手绘草图的图像检索方式越来越受到人们的重视.人们可以通过在触屏上绘制希望检索的图像的草图来表达对图像搜索的特征需求,体现检索中的语义特征含义,从而实现图像信息的检索.然而,由于不同人的绘图方式不同,所绘制的草图或抽象或复杂,使得检索系统难以识别其中的深层语义信息,导致检索准确率低下.为了解决上述问题,本文提出了一种面向草图检索的边缘特征提取算法.首先,提出了一种基于高斯模糊的多尺度边缘提取算法(GBME),通过不断叠加高斯模糊处理后的边缘滤波结果,使得特征边缘获得了更多的图像细节信息,这样更能体现出细节的特征.其次,提出了一种随机采样点混合重心特征描述符(RSB-HOG),通过对图像边缘上的点随机采样后提取 HOG 特征的方式,使自身拥有局部特征描述符特性,并在其特征向量中加入所有采样点的重心方向分布,使其拥有对边缘轮廓信息的把控能力.实验结果表明,在 Flickr15K 数据集上,相比现有技术,上述两点创新均能提高检索准确率,使用上述方法的面向草图的图像检索系统具有更优的检索性能.

关键词: 手绘草图; 图像检索; 特征提取; 特征描述符

中图法分类号: TP391.41

An Edge Feature Extraction Algorithm for Sketch Retrieval

Wang Ruomei^{1,2)}, Jiang Junkun^{1,2)}, Lin Shujin^{2,3)*}, and Wang Fei^{1,2)}

¹⁾ (National Engineering Research Center of Digital Life, Sun Yat-sen University, Guangzhou 511400)

²⁾ (School of Data and Computer Science, Sun Yat-sen University, Guangzhou 511400)

³⁾ (School of Communication and Design, Sun Yat-sen University, Guangzhou 511400)

Abstract: In the face of the sharp increase of image data, how to efficiently retrieve image still is a hot topic for researchers. At present, the mainstream retrieval method is the content-based image retrieval. With the development of touch-screen human-computer interaction technology, people attach more and more importance to the sketch-based image retrieval. People can express the feature requirement of the image by drawing the sketch of the target image on the touch screen, reflecting the semantic feature in retrieval, so as to achieve the retrieval purpose by pictorial information. However, different people have different drawing methods, that makes sketches abstract or complex, making it difficult to identify the deep semantic information in the retrieval system, causing low retrieval accuracy. In order to solve above problems, an edge feature extraction algorithm is proposed. Firstly, a Gaussian blur based multi-scale edge extraction algorithm (GBME) is proposed. By adding the edge filtering results of the Gauss blur processing, the feature edges are

基金项目: 基金名称 1(61432003, 61572531); 基金名称 2(项目编号 1, 项目编号 2, ……); …… 作者 1(出生年—), 性别, 学位, 职称, 若是博导/硕导请注明, 若是 CCF 会员请注明, 主要研究方向为***** (每个研究方向之间用顿号分隔); 作者 2(出生年—), 性别, 学位, 职称, 若是博导/硕导请注明, 若是 CCF 会员请注明, 若是论文通讯作者请注明, 主要研究方向为***** (每个研究方向之间用顿号分隔); ……

obtained more image details, so that the features of image can be reflected more. Secondly, a randomly sampled with barycenter-HOG descriptor (RSB-HOG) is proposed. By random sampling points on the feature edges to extract points' HOG features, RSB-HOG obtains the characteristics of local feature descriptors. And the directional distribution of the barycenter is added to make the descriptor obtain the edge contour information. The experiment shows that, on the Flickr15K dataset, compared with the existing methods, the above two algorithms can improve the retrieval accuracy. The sketch-based image retrieval system using the above method has better retrieval performance.

Key words: sketch; image retrieval; feature extraction; feature descriptor

1 引言

伴随互联网与移动终端的飞速发展, 图像作为信息的主要载体早已融入到了人们生活的方方面面. 数据量的激增, 使人们不得不面临这样一个问题: 如何快速有效地从庞大的图像集中筛选出自己想要的内容. 为解决上述难题, 除采用基于图像内容(图像的颜色、纹理、形状、空间位置关系等特征)进行检索的图像检索技术^[1,2,3] (Content-Based Image Retrieval, 简称 CBIR) 外, 触摸屏人机交互技术的兴起使得基于手绘草图的图像检索技术^[4,5,6,7] (Sketch-Based Image Retrieval, 简称 SBIR) 成为可能: 人们在移动终端的屏幕上绘制目标图像的大致草图, 经过 SBIR 系统的处理, 返回与草图特征近似的系列图像. 该技术使图像检索变得更加方便快捷, 较 CBIR 系统更好地建立起跨越由图像高层抽象语义带来的语义鸿沟的桥梁纽带. 目前已知的最早的 SBIR 系统是由 Kato T 等^[10]提出的, 距今已有 25 余年, 系统检索步骤逐渐归拢至如下形式: 将自然图像转化为与手绘草图近似的中间结果, 即特征边缘, 构成 <自然图像, 特征边缘> 二元映射, 接着提取图像特征以计算手绘草图与特征边缘的相似程度, 得到与手绘草图匹配的特征边缘, 根据映射找到对应的自然图像, 整个流程可大致描述为“手绘草图->特征边缘->自然图像”过程.

手绘草图来源于人, 不同人拥有不同的作图习惯, 人的绘画水平也参差不齐. 如图 1 第一列所示, 对于自然图像“马”, 有的人倾向于绘制马的轮廓, 而有的人则是使用矩形代表马的身体, 用线条代表马的四肢; 有的人拥有更好的绘画基础, 或是接受过相关训练, 绘制的草图拥有马的更多细节(鬃毛、弯曲的四肢), 而有的人会因绘图环境所限(如时间限制), 短时间内绘制的线条过于杂乱. 上述差异将会给系统的匹配带来大量噪声, 例如由矩形与线条构成的马的草图, 与呈矩形的建筑近似, 因此系统极有可能由马的草图检索出方形建

筑物. 对于一幅抽象的草图, 人们能够通过自身积累的经验去推测草图所代表的深层语义, 从而联想到对应的自然图像, 而计算机则需要克服因差异产生的歧义, 抽取鲁棒的特征进行匹配.

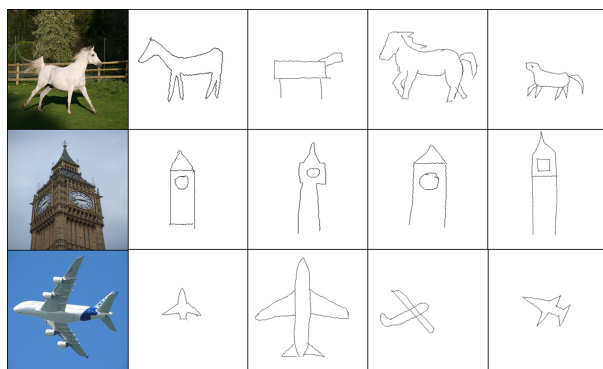


图 1 不同绘制风格的手绘草图

如何解决人们绘制手绘草图所带来的歧义性对图像检索准确率照成的影响, 是 SBIR 系统面临的主要难题, 具体而言主要有如下两个方面:

(1) 特征表示. 将自然图像与手绘草图转换成一种计算机能够识别的数据形式, 这种数据形式称之为特征描述符. 一个有效的特征描述符不仅能够表达自然图像的深层语义信息, 而且能面向手绘草图, 根据草图的笔触方向、线条连贯性等性质, 表达手绘草图的深层语义信息. 系统使用该特征描述符区分特征边缘与草图, 在由绘图差异而产生的歧义与噪声下保持其检索的鲁棒性;

(2) 特征匹配. 根据特征表示方法, 定义合理的度量方式来匹配手绘草图与自然图像的特征. 一个合理的特征匹配方法应是快速有效的, 能够清晰地分辨手绘草图与目标图像的相似之处, 进而得到手绘草图与自然图像的相似度评价;

本文面向草图检索的图像的特征提取问题进行了研究, 主要贡献工作包括: (1) 提出了一种基于高斯模糊的多尺度边缘提取算法 (Gaussian Blur based Multiscale Edge extraction, 简称 GBME), 面对图像信息, GBME 通过不断叠加高斯模糊处理后

的边缘滤波结果,使得特征边缘获得了更多的图像细节信息,这样更能体现出细节的特征;(2)提出了一种随机采样点混合重心特征描述符(Randomly Sampled with Barycentre-HOG,简称RSB-HOG).本文提出的RSB-HOG是全局特征描述符HOG的改进算法,它通过对图像边缘上的点随机采样后提取HOG特征的方式,使自身拥有局部特征描述符特性,并在其特征向量中加入所有采样点的重心方向分布,使其拥有对边缘轮廓信息的把控功能.

实验结果表明,应用本文提出的方法进行面向草图的图像检索,能够拥有更好的检索准确率,与现有的SBIR系统^[10,11,12]相比,在Flickr15数据集下基于本方法的SBIR系统匹配准确率提升了10%有余,所提出的图像特征描述符能够更好地表达手绘草图与自然图像深层语义信息,对绘图差异造成的歧义与噪音问题也有较好的改进.

本文组织结构如下:第一部分介绍SBIR的相关工作,第二部分详细解释RSB-HOG图像特征描述符与GBME边缘提取算法,第三部分是实验结果与分析,第四部分为全文总结.

2 相关工作

近年来,研究人员对SBIR领域关注颇深,提出的特征描述符层出不穷,本节将会介绍几个广泛应用于SBIR系统的相关算法.QVE(Query by Visual Example)作为由Hirata K等^[15]提出的第一个SBIR系统,其与媒体数据库进行特征匹配的理念与系统总体框架的设计思路为后续研究产生了深远的影响.为更好地解决手绘草图与自然图像的语义鸿沟,人们对SBIR系统中的特征表示(即特征描述符的选取)进行了长期的研究,目前主流研究大致将其分为两类:一是全局特征描述符,二是局部特征描述符.

对于全局特征描述符而言,它从全局出发,着重描述图像的整体内容,例如图像的颜色、空间结构信息.由Dalal N等^[16]提出的HOG特征描述符就是其代表之一.HOG作为人体检测算法的图像特征描述符表现优异,其本质是统计图像的梯度方向,突出图像中的纹理与边缘信息.Saavedra J等^[17]将HOG描述符应用于SBIR系统中取得了较好的检索效果,并提出了S-HELO(Soft-Histogram of Edge Local Orientations)特征描述符作为自身工作的创新,S-HELO与HOG的不同之处在于,HOG是对图像所有像素点进行采样,而S-HELO是图像的边缘区域上进行采样.Tu B等^[12]则是使用HOG改进算法GF-HOG(Gradient Field HOG),结合草图的笔画颜色信息对图像进行检索,GF-HOG统计的不是图像中所有像素点的HOG特

征,而是统计构成图像边缘的像素点的HOG特征.Oliva A等^[18]在Friedman A等^[19]研究的基础上提出了GIST特征描述符,GIST用于描述场景特征,将图像从空间域转换到频谱域,定义了5个不同场景所对应的频谱图,在特征匹配环节使用频谱图进行比较.Chalechale A等^[20]提出的ARP(Angular and Radial Partitioning)特征描述符则是按角与径向对图像进行分割,统计分割后扇形区域内的有值像素点的数目,而Zhang Y等^[9]则将HOG与ARP技术结合在一起,提出了AROP(Angle, Radius and Orientation Partition)特征描述符,统计分割后扇形区域内有值像素点的梯度方向,并提出了一种新的图像预处理方法,使得预处理结果与手绘图像更加近似.全局特征描述符拥有较局部特征描述符特征空间更小、匹配速度更快的优点,但由于是“宏观”的描述,所以缺失图像细节,匹配准确率不尽如人意.

局部特征描述符与全局特征描述符的描述方式相反,它是从单张图片的某一特定区域提取特征,着重于图像细节的刻画.Eitz M等^[11]对HOG特征进行了改进,提出了局部特征描述符SHOG(Skethed feature lines in the Histogram of Oriented Gradient),SHOG特征在图像的边缘处提取原始HOG特征,降低了原始HOG的特征空间大小,提高了特征匹配速度.Can Y等^[3]针对边缘信息提出了一种边缘特征匹配算法EI(Edgel Index Algorithm),所含信息量极大,虽从图像匹配准确度上看,与前人研究相比EI占优,但也导致其特征空间数据量极大,不适合应用于海量图像集的场景.Eitz M等^[21]为应对大规模图像检索问题,提出了一种基于度量图像边缘张量的局部特征描述符TENSOR,有效地降低了因图像规模过大而产生的检索时间.

Hu R等^[22]首次有效应用BoF(Bag of Feature,简称BoF)框架进行图像检索,证明了BoF框架能够降低SBIR系统的运算复杂度,该系统视图像由多个视觉词汇组成,视觉词汇由图像特征描述符经聚类后产生,将每张图像的视觉词汇个数当作该图特征.Eitz M等^[11]使用现有的特征描述符(如ShapeContext、SIFT等)与提出的特征描述符(SHOG、SparkFeature),结合BoF框架,建立了SBIR系统,说明了BoF技术的优越性,它使用特征聚类得到的视觉词汇作为最终的图像特征,忽略了人们在绘图时因绘制习惯或者手绘随意性引发的噪声,进而提高了检索准确率,并且该技术降低了图像集的特征空间维度,减少了数据存储消耗.Hu R等^[23]进一步说明,他们提出的使用GF-HOG特征描述符、基于BOF框架的SBIR系统,是具有鲁棒性的,与其他SBIR系统相比,BoF框架提高了图像匹配的准确率,加快了系统检索

速度. Li Y 等^[24]使用 BoF 技术, 提出了一种新型的基于细粒度手绘草图的图像检索系统, 该系统能够识别手绘草图中的细节信息, 并指出通过加入 BoF 技术, 系统的检索准确率得到了进一步提高.

通过以上文献的研究与分析不难发现, 目前研究人员在基于手绘草图的图像检索领域, 更倾向于使用局部特征描述符作为手绘草图的图像特征提取方法, 而不是全局特征描述符, 这是因为相比自然图像, 手绘草图更多由白色背景与黑色线条组成, 拥有极少的颜色信息与空间结构信息, 从“宏观”上描述并不能良好表示草图的深层语义. 但使用局部特征描述符将会引入绘图噪音、加大特征空间数据量, 因此研究人员使用 BoF 技术对特征进行后处理, 忽略因噪音产生的影响, 提高系统检索准确率, 并且降低特征空间, 加快检索速度. 上述分析为本文的研究方法提供了思路.

本文的研究工作包括两个部分: 1) 提出了一种基于高斯模糊的多尺度边缘提取算法(Gaussian Blur based Multiscale Edge extraction, 简称 GBME); 2) 提出了一种随机采样点混合重心特征描述符(Randomly Sampled with Barycentre-HOG, 简称 RSB-HOG). 最后应用本文提出的算法建立了基于手绘草图的图像检索系统, 验证了本文提出的算法的优越性.

3 面向草图检索的边缘特征提取算法

3.1 基于高斯模糊的多尺度边缘提取算法

SBIR 系统通过提取自然图像的特征边缘, 将特征边缘看作是自然图像与手绘草图的中间结果, 比较特征边缘与手绘草图二者的相似度, 因此若特征边缘越能表达自然图像的深层语义信息, 系统的检索结果将会更加准确. 因此, 问题的关键是将图像边缘特征与手绘草图的特征更加相似, 针对这个问题, 本文提出了基于高斯模糊的多尺度边缘提取算法 GBME(Gaussian Blur based Multiscale Edge extraction).

如图 2 所示, 图 2(a)为“伦敦塔”主题的自然图像子集, 图 2(b)为直接使用 Canny 算子得到的特征边缘图像, 图 2(c)为 GBME 算法产生的特征边缘图像, 图 2(d)为与自然图像“伦敦塔”相匹配的手绘草图, 图 2(e)为非“伦敦塔”的其它手绘草图. 显然, 图 2(c)较图 2(b)拥有更多的图像细节, 并且, 比较图 2(b)与图 2(e)、图 2(c)与图 2(d), 图 2(b)近似为伦敦塔的建筑轮廓, 更易与图 2(e)中各种尖形物体轮廓的手绘草图相匹配, 而不会与图 2(d)含有内部结构的草图相匹配. 因此, 原算法中直接使用 Canny 算子提取边缘, 极大可能导致匹配错误, 降低系统的检索准确率, 而 GBME 算法因为其丰富

的细节, 避免了错误匹配问题的出现.

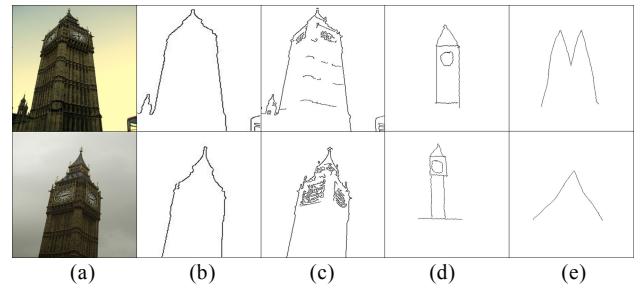


图 2 两种特征边缘提取算法对比图

为了能够更多的获得图像的边缘特征, 本文提出的 GBME 算法采用迭代的形式, 不断从高斯模糊处理后的多尺度图像中运用边缘算子提取特征边缘, 随后对迭代产生的特征边缘集取和, 最终得到含有更多细节的图像特征边缘.

不同于已有 SBIR 系统^[8,11,12,13]直接使用边缘算子(如 Canny、Laplace、Sobel)提取特征边缘, 本算法意在通过使用高斯模糊以此模拟人眼的特性: 文献[14]指出, 高斯型带通滤波器为模拟人眼视觉的典型模型之一. 本算法使用不同高斯核大小对图像进行高斯模糊操作, 得到不同尺度下的图像, 近似模仿人类在接收到视觉信号后由模糊到清晰的这一人眼对焦过程, 使得提取出来的边缘更加接近人脑所能感知的结果, 从直观上看, 由于提取了不同尺度下的特征边缘, 使得结果拥有更多的边缘细节. 图像特征边缘提取算法 GBME 的详细描述如下所示:

输入. 灰度图 Img .

输出. 灰度图的特征边缘 $Edge$.

Step1. 读入图像 Img .

Step2. 设置迭代次数 S .

Step2.1. 当 S 大于 0, 执行如下步骤:

Step2.2. $Img'' = \text{GaussianBlur}(Img', S)$

Step2.3. $Edge = Edge + \text{Canny}(Img'')$

3.2 随机采样点混合重心特征描述符

在提取图像深层语义中, 本文提出了一种随机采样点混合重心图像特征描述符 RSB-HOG(Randomly Sampled with Barycenter-HOG), 它是全局特征描述符 HOG^[16]的改进算法. 如图 3 所示, RSB-HOG 特征描述符在特征边缘处进行随机采样, 采样点与重心分别以红点与蓝点示意, 统计红框处特征边缘的梯度方向作为采样点(红点)的 HOG 特征, 并统计上述采样点的重心(蓝点)至各个采样点的方向分布, 构成方向分布直方图作为对采样点特征的补充. 因此, 若记 RSB-HOG 的特征空间为 G , 随机采样点数目为 N , 单个随机采样点的特征向量为 f_i , 构成随机采样点

的特征矩阵为 F , 重心点的特征向量为 B , 则 G 可由公式(1)表示.

RSB-HOG features extraction

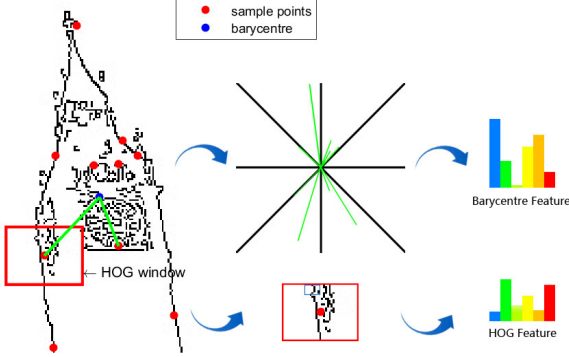


图3 RSB-HOG 特征描述符示意图

对于图像边缘的定义, 由于特征提取步的输入是手绘草图与图像边缘集, 其本质是二值图像, 因此如果输入图像中的任意像素点拥有灰度值, 该点即可定义为图像边缘上的点. 如式(2)所示: 图像中的任意像素点 (x, y) 上的灰度值为 $gray(x, y)$, 若是边缘点, 其边缘属性 $IsEdge(x, y)$ 定义为 1, 否则为 0.

$$F = [f_1, f_2, f_3, \dots, f_N] \quad (1)$$

$$G = [F, B]$$

$$IsEdge = \begin{cases} 1, & \text{if } gray(x, y) > 0 \\ 0, & \text{others} \end{cases} \quad (2)$$

提取 RSB-HOG 特征的详细步骤说明如下:

- 1) 对输入图像的边缘点集进行随机采样, 得到采样点集合 S , 记其点数为 β ;
- 2) 计算采样点的 HOG 特征向量 f . 对于采样点 $s(s \in S)$, 该点在输入图像上的坐标为 (x, y) , 其灰度值为 $gray(x, y)$:

- a) 求图像中所有像素点的灰度梯度;
- b) 对于采样点 S , 设置窗口 W . 其中 W 的大小为 $w = M * N$, 中心为 $s(x, y)$;
- c) 在窗口 W 中设置滑块 C , 其大小为 $c = m * n$, 滑块 C 在窗口 W 中从上至下、从左至右依次滑动给定步长 (α_x, α_y) 的距离. 每滑动一步, 记当前位置的滑块为 C_i , 滑块总共可移动次数为 λ (如式(3)所示), 即 $i \in \{1, 2, \dots, \lambda\}$;

- d) 将 C_i 中所有点的梯度方向进行投影, 使用

直方图 $hist_s$ 统计每点的梯度方向 $grad$ 命中标定方向(如式(4)所示)的次数;

- e) 直方图 $hist_s$ 归一化, 结果即为采样点 S 的 HOG 特征向量 f , 其特征维度为 λ .

- 3) 计算所有采样点的重心 $b = \sum s / \beta$;

- 4) 将重心 b 至每个采样点 S 的方向进行投影, 使用直方图 $hist_b$ 统计每个方向 $dirc$ 命中标定方向 (如式(5)所示) 的次数;

- 5) 直方图 $hist_b$ 归一化, 结果即为 B , 与上步特

征 F 结合起来构成 RSB-HOG 特征 $G = [F, B]$.

本特征描述符 RSB-HOG 设置的采样点数目为 500, 采样窗口大小设置为 $256 * 256$, 滑块大小设置为 $80 * 60$.

$$\lambda = \left\lceil \frac{M - m}{\alpha_x} \right\rceil * \left\lceil \frac{N - n}{\alpha_y} \right\rceil \quad (3)$$

$$hist_s(k) = \begin{cases} hist_s(k) + 1, & \text{if } dirc \in \left[\frac{\pi}{6}(k-1), \frac{\pi}{6}k \right] \\ hist_s(k), & \text{others} \end{cases} \quad (4)$$

$$k = 1, 2, \dots, 12$$

$$hist_b(k) = \begin{cases} hist_b(k) + 1, & \text{if } grad \in \left[\frac{\pi}{6}(k-1), \frac{\pi}{6}k \right] \\ hist_b(k), & \text{others} \end{cases} \quad (5)$$

$$k = 1, 2, \dots, 12$$

为减少因采用随机采样点提取特征的策略而带来的图像语义扰动, 降低图像特征空间, 本文采用了 BoF 技术. 本技术采用聚类算法对每幅图像的局部特征进行聚类, 得到的类心定义为视觉字典 (visual codebook) 中的视觉词汇 (visual words), 图像中的每个特征都会被映射到视觉字典中的某个词上, 因此可将每幅图像描述为多个视觉词汇的无序集合, 即图像视觉词汇表, 表中所含视觉词汇数目反映了视觉词汇表的大小. 图像特征的匹配则转化成图像视觉词汇表的相似度比较, 相比图像的高维特征空间, 视觉词汇表是一个维数由视觉字典确定的直方图向量, 其耗费空间更小. 具体步骤如下所述:

- 1) 使用 k-means++ 算法对所有输入图像的特征集

- 合进行聚类操作,产生的类心即为视觉词汇,类心数量即为该视觉字典中含有的视觉词汇总数量;
- 2) 将每张图像的特征点映射至上步求得的视觉词汇上,统计每张图像中视觉词汇出现的次数,以此构建该图的视觉词汇表,由与图像一一对应的视觉词汇表构成表集;
- 3) 对于手绘草图,与 2 步相同,构建手绘草图的视觉词汇表;
- 4) 使用 tf-idf 算法^[24],对上述两种表集进行加权运算,得到最终的视觉词汇表。

3.3 相似度匹配

在相似度匹配模块中,输入是上个模块(深层语义提取模块)的运算结果,包含手绘草图的视觉词汇表与图像集的视觉词汇矩阵,分别记作 vb_{sketch} 与 VB_{img} . 视觉词汇矩阵 VB_{img} 中的一行代表图像集中的一张照片的视觉词汇表,即 $vb_{img}^i \in VB_{img}$. 手绘草图视觉词汇表 vb_{sketch} 与图像视觉词汇表 vb_{img}^i 两者的维度即为 BoF 中设定的视觉词汇数目,记数目为 λ , 图像集中的自然图像数目为 Φ .

本文尝试了多种相似度计算方法,包括余弦相似度(记作 COS, 如式(6)所示)、均方根距离、平均绝对距离、直方图相交距离. 经过实验分析后发现,使用 COS 作为相似度计算方法,系统的检索准确率最高.

$$COS = \frac{\sum_{i=1}^{\Phi} vb_{sketch} \times vb_{img}^i}{\sqrt{\left[\sum_{i=1}^{\Phi} (vb_{sketch})^2 \right] \times \left[\sum_{i=1}^{\Phi} (vb_{img}^i)^2 \right]}} \quad (6)$$

4 实验结果与分析

为了验证本文提出的算法进行了实验对比,本实验通过图像检索准确率来比较所提方法与目前相关方法的优劣,同时研究了在不同参数下对本文所提方法准确率的影响. 为使实验结论具有普遍性和代表性,本文采用真实图像数据集 Flickr15K^[23]作为实验数据,多篇论文^[11,12,13,23,24]均使用该数据集作为衡量文中 SBIR 系统的检索效果的评价标准,并使用平均准确率(Mean Average Precision, 简称 MAP)与精确率-召回率曲线(Precision-Recall, 简称 PR 曲线)作为系统检索效果评价指标.

本节实验均是在 Intel Core i7-4970 3.6GHZ 处理器、16GB 内存、windows 10 64 位操作系统的台式机以及 Microsoft visual studio 2015 开发环境下完成的.

4.1 基于草图绘制的图像检索框架

为验证本文提出的 RSB-HOG 图像特征描述符与 GBME 边缘提取算法对于 SBIR 系统而言是有效、可靠的,能够良好地建立起跨越由图像高层抽象语义带来的语义鸿沟的桥梁纽带,本文根据 RSB-HOG 图像特征描述符与 GBME 边缘提取算法建立了 SBIR 系统,其框架如图 4 所示,主要包括以下三部分:(1)特征边缘提取模块中,本文根据基于高斯模糊的多尺度图像边缘提取算法 GBME,不断迭代模糊处理后的边缘滤波结果,得到更加丰富的图像特征边缘;(2)深层语义提取模块中,本文采用随机采样点混合重心图像特征描述符 RSB-HOG 提取手绘草图与图像特征边缘的局部特征,该特征能够较好地克服图像语义鸿沟带来的歧义,后使用 BoF 技术对其特征空间进行降维操作,分别构建手绘草图与图像集的视觉词汇表,完成对上述两者的特征提取;(3)相似度匹配模块中,使用上一模块得到的两个视觉词汇表进行相似度计算,其结果视作该次检索请求中的手绘草图与图像集的相似度,对其排序后,有序图像集的前 k 张图像子集即为本次请求的检索结果. 经过多次实验对比,本文发现 GBME 算法中迭代次数设置为 17、BoF 中视觉字典大小设置为 3000,其系统检索效果最优.

4.2 实验数据集

数据集 Flickr15K 分为图像集、草图集和参考标注(ground-truth)这三个部分,图 5 展示了数据集的一些例子. 其中,图像集中含有从 Flickr 图片分享网站中随机爬取的 14660 张自然图像,所有图像被分为例如“pyramid”、“bird”、“heart_shape”等 60 个基础类别,平均每个类别含有 200 余张图片;草图集含有由 10 位未经过专业训练的普通人针对 33 类物品所绘制的 330 张手绘草图;参考标注文件里含有每张草图所对应的图像编号,例如 1 号实验者绘制的第一张手绘草图为简单圆形,该张草图应能检索到含有“fire_balloon”、“London_eye”、“moon”等类别的图像,因此该张手绘草图在参考标注文件里就拥有相应的自然图像的文件名标注,上述内容均经过人工标注以保证其类别的准确性.

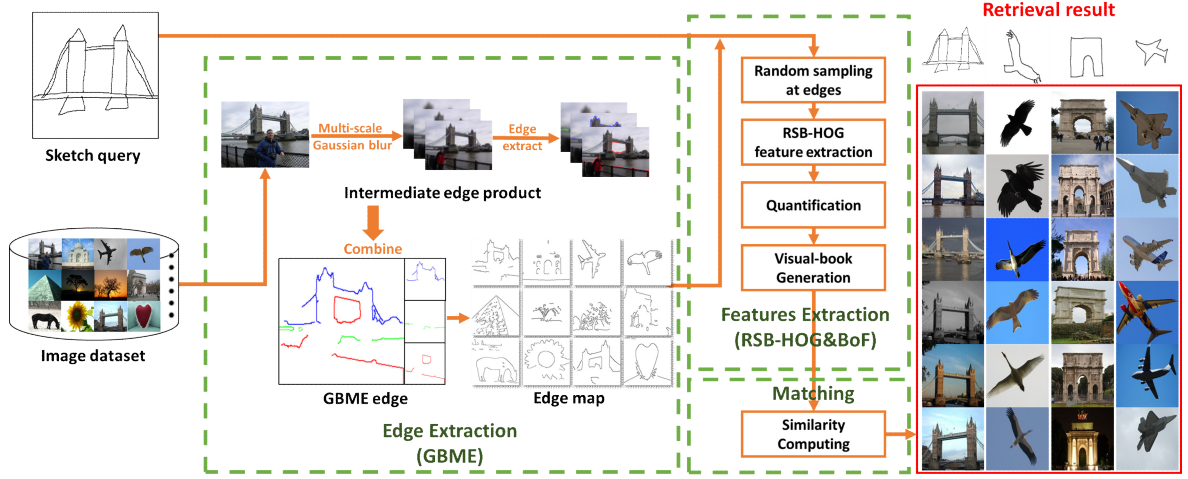


图 4 基于本文方法的 SBIR 系统框架图

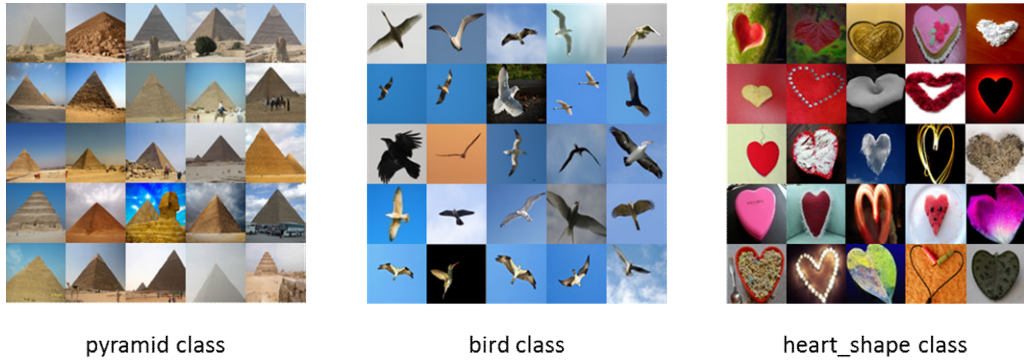


图 5 Flickr15K 中的部分自然图像

4.3 检索效果评价

为客观评价本文提出的 SBIR 系统的检索效果, 本文对比了在使用相同数据集 Flickr15K 下的多个 SBIR 系统: SHOG-BOF^[11]、GFHOG-BOF^[12]、InvertedIndex^[12]、CDMR-BF-fGALIF^[13]、RSB-HOG-BOF(本文方法)。为使对比简单有效, 上述系统均采用各自文中描述的最佳参数进行设置, 以得到各个系统最好的检索结果。另外, 为检验本文提出的图像边缘提取算法的有效性 with 泛用性, 本文还设计了另一组对比实验, 对比样本为使用本文提出的边缘提取算法的 SHOG-BOF 系统(记为 SHOG-BOF-EDGE1)、InvertedIndex 系统(记为 InvertedIndex-EDGE1), 与使用原始边缘提取算法的 SHOG-BOF 系统(记为 SHOG-BOF-EDGE2)、InvertedIndex 系统(记为 InvertedIndex-EDGE2)。

本文采用 MAP 与 PR 曲线作为检索效果评价指标。首先, 对于本文采用的图像边缘提取算法 GBME, 为使对比客观合理, 本文采用在相同数据集下使用相同特征描述符、不同边缘提取方式的 SBIR 系统进行检索实验, 评价标准使用 PR 曲线与 MAP 值, 实验结果见图 6 与表 1。在图 6(a)中, 使用 GBME 图像边缘提取算法、特征描述符为 SHOG

的 SBIR 系统 SHOG-BOF-EDGE1 的曲线明显优于使用原始提取算法的 SHOG-BOF-EDGE2。从图像上看, 在召回区间 $[0, 0.1]$ 处, SHOG-BOF-EDGE1 的精确度下降更加平缓。在表 2 中, SHOG-BOF-EDGE1 的 MAP 值大于 SHOG-BOF-EDGE2 的 MAP 值, 前者较后者提高了 24.50%。而在图 6(b)中, 使用特征描述符 InvertedIndex 的 SBIR 系统中, 在召回区间 $[0, 0.15]$ 处, InvertedIndex-EDGE1 与 InvertedIndex-EDGE2 的精确度相当, 随后前者下降较后者缓慢, 在召回率为 0.5 时, 两者精确度相差最大。就 MAP 值而言, 前者较后者提高了 20.63%。

从实验结果中我们可以得到如下结论, 本文提出的图像边缘提取算法 GBME, 具有一般性, 能够提升多个 SBIR 系统的检索准确度, 也证实了在基于手绘草图的图像检索研究中, 若图像边缘含有自然图像的更多信息, 系统进行特征匹配的效果将会更好。

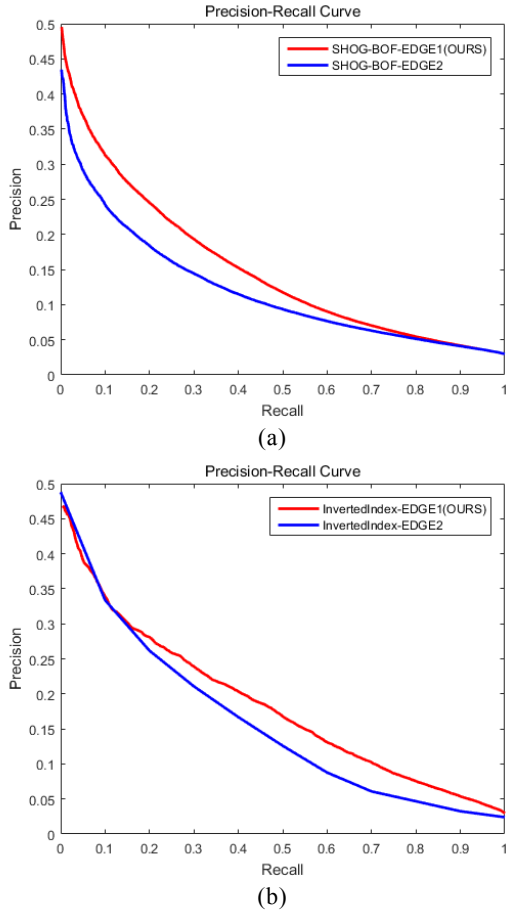


图 6 在不同边缘提取算法下的系统 PR 曲线图对比

表 1 不同图像边缘提取算法下的 MAP 值对比

方法	MAP
SHOG-BOF-EDGE1	19.31%
SHOG-BOF-EDGE2	15.51%
InvertedIndex-EDGE1	20.64%
InvertedIndex-EDGE2	18.21%

接着, 图 7 展示了 RSB-HOG-BOF、InvertedIndex、SHOG-BOF、CDMR-BF-fGALIF、GFHOG-BOF 的 PR 曲线, 可以看到, 本文提出的 RSB-HOG-BOF 系统在不同召回率下, 除 CDMR-BF-fGALIF 外, 其精确率均高于其余系统。GFHOG-BOF 在召回区间 $[0, 0.02]$ 处, 其准确率陡降严重; SHOG-BOF 在召回区间 $[0, 0.1]$ 处, 其准确率也存在一定程度的下降过快的情况; 反观 RSB-HOG-BOF, 较其它系统, 随着召回率的提高, 准确率下降平缓; 在召回区间 $[0.4, 1]$ 处, CDMR-BF-fGALIF 的精确率反超其余四种方法。

本实验计算了每个方法的 MAP 值, 作为评价每种方法对图像分类能力的指标, 如表 2 所示。从表中我们可以看出, 在相同图像集与手绘草图检索请求下, 本文提出的方法 RSB-HOG-BOF 的

MAP 值较其余方法最高, 比近三年内提出的方法 CDMR-BF-fGALIF 提高了 14%, 方法 InvertedIndex 提高了 40.86%。而 MAP 值最低的方法是 GFHOG-BOF, 其原因在于 GF-HOG 特征描述符是全局特征描述符, 通过计算图像边缘的张量、求解张量稀疏矩阵的方式作为图像特征, 经过 BoF 技术处理后, 丢失了过多图像细节信息。

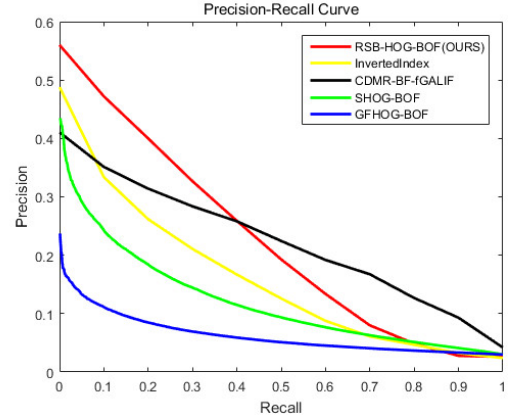


图 7 多种方法的 PR 曲线图对比

表 2 多种方法的 MAP 值对比

方法	MAP
RSB-HOG-BOF(OURS)	25.85%
CDMR-BF-fGALIF	22.50%
InvertedIndex	18.21%
SHOG-BOF	15.51%
GFHOG-BOF	8.27%

5 结 语

为构建上述 SBIR 系统, 本文首先在图像预处理环节, 使用图像边缘提取算法 GBME 以得到更多的图像边缘信息, 使“手绘草图-特征边缘-自然图像”三者的联系更加紧密, 接着提出一种图像特征描述符 RSB-HOG, 意在将 HOG 特征局部化, 并添加采样点重心方向信息以增强图像特征细节与图像轮廓的辨别能力, 随后使用 BoF 框架将图像特征转化成视觉词汇表, 降低特征空间维度, 加快特征匹配速度, 最后选择余弦相似度作为相似度计算方法, SBIR 系统返回与手绘草图相似度最大的前 k 张自然图像。从实验结果的 PR 曲线可以看出, 本系统在数据集 Flickr15K 上具有良好的检索性能, 并且优于现有的多个 SBIR 系统。其中, 本文提出的特征描述子 RSB-HOG 能够较为完整地表达图像信息, 提高匹配的准确率, 而自然图像边缘提取算法 GBME 对于 SBIR 系统具有通用性, 对于不同图像特征描述符, 该算法均改善了系统的检索效果。

由于在检索环节中, 本文使用了 BoF 框架进行特征降维处理, BoF 框架在聚类构建视觉词汇的时候, 使用的是 k-means++ 聚类算法, 因此具有一定的随机性, 导致实验对比存在一定的误差, 重复实验无法得到相同结果。另外, SBIR 系统的检索理论, 是建立在手绘草图从外观上能够近似等同于自然图片的假设之上, 而一般人类并没有接受过绘画训练, 因此草草几笔, 虽然人类能够通过生活经验辨识草图, 但计算机会因为其手绘草图过于简单而出现误分类的情况。

参考文献(References):

- [1] Guo J M, Prasetyo H. Content-based image retrieval using features extracted from halftoning-based block truncation coding.[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 24(3):1010-24.
- [2] Yang X, Qian X, Xue Y. Scalable mobile image retrieval by exploring contextual saliency[J]. IEEE Transactions on Image Processing, 2015, 24(6):1709.
- [6] Yang Cao, Changhu Wang, Liqing Zhang, et al. Edgel index for large-scale sketch-based image search[C]// Computer Vision and Pattern Recognition. IEEE, 2011:761-768.
- [7] Chalechale A, Mertins A, Naghdy G. Edge image description using angular radial partitioning[J]. IEE Proceedings - Vision Image and Signal Processing, 2004, 151(2):93-101.
- [8] Eitz M, Hays J, Alexa M. How do humans sketch objects?[J]. Acm Transactions on Graphics, 2012, 31(4):1-10.
- [9] Zhang Y, Qian X, Tan X, et al. Sketch-Based Image Retrieval by Salient Contour Reinforcement[J]. IEEE Transactions on Multimedia, 2016, 18(8):1604-1615.
- [10] Kato T, Kurita T, Otsu N, et al. A sketch retrieval method for full color image database-query by visual example[C]// Conference A : Computer Vision and Applications, International Conference on Pattern Recognition. 1992:530-533.
- [11] Eitz M, Hildebrand K, Boubekeur T, et al. Sketch-Based Image Retrieval: Benchmark and Bag-of-Features Descriptors[J]. IEEE Transactions on Visualization & Computer Graphics, 2011, 17(11):1624-1636.
- [12] Tu B, Collomosse J. Scalable Sketch-Based Image Retrieval Using Color Gradient Features[C]// IEEE International Conference on Computer Vision Workshop. IEEE Computer Society, 2015:1012-1019.
- [13] Furuya T, Ohbuchi R. Visual Saliency Weighting and Cross-Domain Manifold Ranking for Sketch-Based Image Retrieval[M]// MultiMedia Modeling. Springer International Publishing, 2014:37-49.
- [14] Lloyd J M. Thermal Imaging Systems[M]. Springer US, 1975.
- [15] Hirata K, Kato T. Query by Visual Example - Content based Image Retrieval[C]// International Conference on Extending Database Technology: Advances in Database Technology. Springer-Verlag, 1992:56-71.
- [16] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005:886-893.
- [17] Saavedra J M. Sketch based image retrieval using a soft computation of the histogram of edge local orientations (S-HELO)[C]// IEEE International Conference on Image Processing. IEEE, 2015:2998-3002.
- [18] Oliva A, Torralba A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope[J]. International Journal of Computer Vision, 2001, 42(3):145-175.
- [19] Friedman A. Framing pictures: the role of knowledge in automatized encoding and memory for gist.[J]. J Exp Psychol Gen, 1979, 108(3):316-355.
- [20] Chalechale A, Mertins A, Naghdy G. Edge image description using angular radial partitioning[J]. IEE Proceedings - Vision Image and Signal Processing, 2004, 151(2):93-101.
- [21] Eitz M, Hildebrand K, Boubekeur T, et al. A descriptor for large scale image retrieval based on sketched feature lines[C]// Eurographics Symposium on Sketch-Based Interfaces and Modeling. ACM, 2009:29-36.
- [22] Hu R, Barnard M, Collomosse J. Gradient field descriptor for sketch based retrieval and localization[C]// IEEE International Conference on Image Processing. IEEE, 2010:1025-1028.
- [23] Hu R, Collomosse J. A performance evaluation of gradient field HOG descriptor for sketch based image retrieval[J]. Computer Vision & Image Understanding, 2013, 117(7):790-806.
- [24] Li Y, Hospedales T, Song Y Z, et al. Intra-category sketch-based image retrieval by matching deformable part models[C]// British Machine Vision Conference. 2014:115.1-115.12.
- [25] Sivic J, Zisserman A. Video Google: A Text Retrieval Approach to Object Matching in Videos[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2003:1470.

附中文参考文献:

- [3] 王涛, 胡事民, 孙家广. 基于颜色—空间特征的图像检索[J]. 软件学报, 2002, 13(10):2031-2036.
- [4] 钱晶, 汪增福, 曹洋. 基于草图查询的语义对象图像检索[J]. 小型微型计算机系统, 2010, 31(10):2079-2087.
- [5] 李曼舞, 孙正兴, 顾庆东. 基于轮廓的手绘草图检索研究[J]. 计算机应用, 2005, 25(2):344-347.