

# 融合相机元信息的基于区域的手机图片搜索

马桔 路红 郭跃飞

(复旦大学计算科学与工程系, 上海 200433)

**摘要** 照相手机的流行及其具有的随身携带性,使得人们能够随时随地进行拍照。相比传统的相机来说,手机的联网性使得人们能够实时地进行图片搜索和分享。由此而来的手机图片数量的急剧增长,又使得如何高效地组织、管理以及检索这些图片成为了研究热点。为了高效地进行手机图片搜索,提出了一种融合相机元信息(Exif)的基于区域的手机图片搜索算法,同时利用这个算法实现了一个在线的用户手机拍摄图片的搜索系统。通过与传统的基于内容的手机图片搜索的对比可见,该算法通过融合进相机元信息以及物体的区域特征,在一定程度上降低了“语义鸿沟”问题。实验结果表明,该算法优于传统的基于底层特征的图片搜索算法。

**关键词** 图片搜索 基于内容的图片搜索 基于区域的方法 照相手机 Exif

**中图分类号**: TP301.6 **文献标识码**: A **文章编号**: 1006-8961(2007)10-1766-04

## Region Based Cameraphone Image Retrieval Incorporated with Metadata

MA Jie, LU Hong, GUO Yue-fei

(Department of Computer Science & Engineering Fudan University, Shanghai 200433)

**Abstract** The convenient attribute of cameraphone makes it popular nowadays; people can use it to snap picture at any time and everywhere. At the same time with the advantage of inherent network connectivity, we can share the snapped image with others immediately or search for similar images from the web. The above advantages of cameraphone make the quantity of cameraphone images boosting quickly, which results in the problem of how to efficient organize, manage and retrieve. In this paper, a new region based cameraphone images retrieval incorporated with metadata method is proposed to improve the image retrieval performance. Using this algorithm, we implemented an online demo system to allow users to upload their snapped images and do searching similar images. Compared to traditional contented based image retrieval (CBIR) method, our method uses the region feature of the image incorporated with metadata of cameraphone to bridge the semantic gap between the low-level feature of images and semantics. The experimental results show higher efficiency of our methods than the traditional CBIR method.

**Keywords** image retrieval, contented based image retrieval(CBIR), region based method, cameraphone, Exif

## 1 引言

随着照相手机的性价比的不断提高,其销售数量已经跟传统的手机不相上下。同时照相手机与传统的数码相机相比较又具有如下一些优势:(1)当利用照相手机进行拍照时,能够同时获得一些相机元信息和用户联系人信息(拍照时间、地点、日历和

联系人信息等);(2)由于照相手机具有内在的联网性,使得照相手机用户能够随时地分享图片或者进行图片搜索;(3)智能照相手机拥有一个开放的可编程接口(如,J2ME、Symbian、BREW,等),它使得能够针对应用进行二次开发<sup>[1]</sup>。上述的优势使得照相手机图片数量急剧增长,因此如何高效地组织、管理和搜索图片的需求应运而生。传统的基于内容的图片搜索是利用提取图片的底层特征的方法来进

基金项目:国家自然科学基金项目(60533100,60402007);上海市研究与发展基金项目(05QM1403,065115017,06DZ15008)

收稿日期:2006-06-08;改回日期:2007-07-02

第一作者简介:马桔(1982-),男,复旦大学计算机科学与工程系硕士研究生。主要研究方向为图像视频高层语义提取、分类、检索、模式识别等。E-mail: 052021181@fudan.edu.cn

行相似度查询。基于内容的图片搜索在20世纪90年代初被提出来,为了克服基于手工标注的图片搜索的两大难题——图片数量的巨大性以及手工标注的主观性,初期,基于内容的图片搜索获得了相当的成功,产生了许多研究和商业的应用系统,其中卡耐基梅隆大学开发的 Informedia 系统<sup>[2]</sup>以及 IBM 公司开发的 QBIC 系统<sup>[3]</sup>即为两个典型的系统。

尽管基于内容的图片搜索解决了手工标注的问题,但是其本身同样存在图像的底层特征和图片之间的语义鸿沟问题。为了解决语义鸿沟问题,人们进行了许多研究。Liu 等概括出了以下3种主要的解决语义鸿沟的方法<sup>[4]</sup>:(1)基于区域的图片搜索;(2)在搜索中引入相关反馈;(3)对高层语义进行建模。文献[4]利用方法3对高层语义进行建模,并将 HSV 颜色空间映射成93个语义颜色名字,如(0.36,0.99,0.65)表示绿色的草,而(0.62,0.67,0.92)表示蓝天。但此方法是相对粗糙的和精确的,而且也容易受到光照等影响。

上述图像分析检索的方法都只考虑了图像中的像素,而忽略了图像本身所携带的元信息。在图像中和图像之间存在着一定的上下文关系,这种信息

能够被用来进行图像分析。如 Boutell 等利用相机元信息以及图片的底层特征进行语义场景分类<sup>[5]</sup>; Luo 等概括出了图像中存在的时间、空间和图片元信息的上下文关系<sup>[6]</sup>。其中相机元信息是指 Exif 信息<sup>[7]</sup>,例如一组标准的相机元信息包含在 JPEG 图片中,Exif 信息记录着拍摄时的相机信息,如焦距长度、曝光时间、光圈  $F$  值、闪光灯、物理距离和曝光偏差等。文献[5]利用这些信息进行图像的语义分析。如闪光灯在室内使用的频率远大于室外;光圈  $F$  值越大,图像越是室外的风景,而不是具体的物体。本文利用基于区域的图像底层特征以及相机的元信息来提高图片的搜索性能。

## 2 在线手机图片搜索系统框架

本文所实现的系统包括如下部分:移动手机端、服务器端、电脑用户端。如图1所示,其中服务器端算法引擎是本文研究的重点,其包含如下部分:图像分割、提取基于区域的底层特征、特征相似性计算和排序,然后通过整合图片的 Exif 信息来调整图片的相似度。

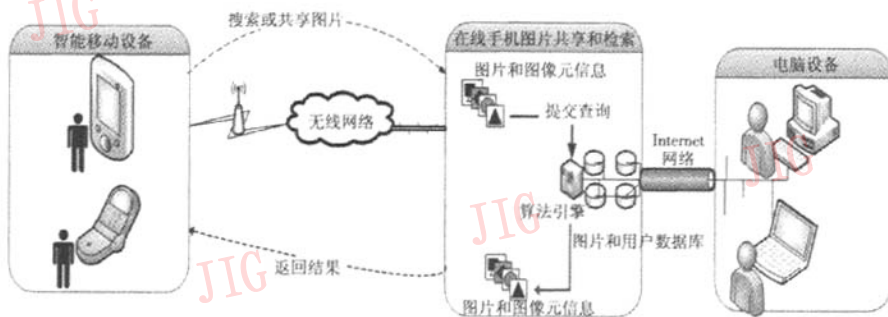


图1 在线手机图片搜索系统

Fig.1 Online cameraphone image retrieval system

### 2.1 图像分割以及底层特征提取

图像分割同样也是计算机视觉中的研究重点,目前已经提出了很多的分割算法<sup>[8,9]</sup>,通过在数据集上的实验发现,JSEG 算法<sup>[8]</sup>对于图片的分割效果较好。本实验先采用该分割算法进行图像分割,而参数的设置则根据文献[8]进行图像分割后,可忽略那些分割出的小面积的区域(占图像总面积小于10%);然后在显著区域上进行特征提取,并使用了 MPEG-7<sup>[10]</sup>中的颜色布局描述子(color layout descriptor)和可伸展颜色描述子(scalable color

descriptor)来描述区域的特征。

### 2.2 图片的相似度度量和排序

给定一张查询图片,经过图片分割和底层特征提取之后,可把图片描述成如下形式:

$$I_i = \{(f_j, w_j), \dots, (f_n, w_n)\} \quad (1)$$

其中,  $(f_j, w_j)$  描述了第  $i$  幅图片的第  $j$  个区域特征向量,  $f_j$  表述第  $j$  个区域的底层特征向量,  $w_j$  描述了第  $j$  个区域的权重,其由式(2)所定义。

$$w_j = \text{size\_of(region\_j)} / \text{size\_of(image\_i)} \quad (2)$$

另外,可以同样的方式将数据库中的图片也描述成

式(1)的形式,如

查询第  $i$  幅图片

$Q_i = \{(f_1, w_1), \dots, (f_n, w_n)\}$ ;  $n$  个区域源第  $t$  幅源图片

$Y_t = \{(f_1, w_1), \dots, (f_m, w_m)\}$ ;  $m$  个区域

由此就可以将图像之间的相似性转化成计算不同区域个数的  $Q_i$  和  $Y_t$  之间的相似性。本文采用 EMD 算法<sup>[11]</sup>来计算二者的相似性。EMD 算法的原理是:将  $Q_i$  和  $Y_t$  分别认为是散落在地球上不同地方的大小不一、数目不等的洞和土堆。算法的目标是利用这些土堆来尽可能地填满这些洞,同时保持总的运输距离最小。在 EMD 算法中,一个洞和一堆土之间的距离被称作为基准距离,用户可以自己选择定义该距离,如一阶范式,或者欧氏距离等。本文使用了欧氏距离(如式(3)所示,其中,  $d$  为底层特征的维数),由于其具有易计算性和良好的图片搜索效果。

$$D(f_j^Q, f_k^Y) = \sqrt{(f_{j,1}^Q - f_{k,1}^Y)^2 + (f_{j,2}^Q - f_{k,2}^Y)^2 + \dots + (f_{j,d}^Q - f_{k,d}^Y)^2} \quad (3)$$

因此,两幅图片的 EMD 距离计算如式(4)所示:

$$EMD(Q_i, Y_t) = \frac{\sum_{m \in Q_i} \sum_{n \in Y_t} c_{m,n} D(f_m^Q, f_n^Y)}{\sum_{m \in Q_i} \sum_{n \in Y_t} c_{m,n}} \quad (4)$$

其中,  $c_{m,n}$  表述第  $i$  幅查询图片  $Q_i$  的第  $m$  区域和第  $t$  幅源图片  $Y_t$  的第  $n$  区域之间的路径。函数  $D$  通过式(3)定义,分母则为归一化因子。EMD 算法是通过迭代的方式最小化式(4)来找到最优的图片  $Q_i$  和  $Y_t$  之间的路径  $c_{m,n}$ 。最后再根据 EMD 值从小到大进行排序。

### 2.3 通过 Exif 信息调整图片的相似性排序

除了利用图片的像素值,同时也利用了图片的 Exif 信息来调整图片之间的相似度。对于不同大小的物体,光圈  $F$  值是不同的,曝光时间也是随着物体本身颜色的明亮程度而不同。从统计的角度通过分析数据集发现,光圈  $F$  值对于汽车图片来说比较小,而对于高楼来说,则曝光时间比较小(如表 1 所示)。

根据上述结果,本文使用不同的曝光时间(exposure time, ET)以及光圈  $F$  值( $F$  number, FN)来调整搜索的结果。同样使用欧氏距离来度量两幅图像的 Exif 信息之间的差异为

$$S_E = D_{ET}(q, t_i) + D_{FN}(q, t_i) \quad (5)$$

利用 Exif 信息调整后的最终的查询图片  $q$  和第  $i$  幅

表 1 6 个概念的曝光时间以及光圈  $F$  值的均值

Tab.1 Mean of exposure time and mean of  $F$ -number of six concepts

概念	平均曝光时间(s)	平均光圈 $F$ 值
汽车	0.007 254 66	3.318 52
园林	0.005 205 88	4.364 71
草地	0.004 967 21	5.186 89
高楼	0.001 905 69	5.6
雕塑	0.004 025 76	5.6
红色楼房	0.003 545 69	5.492 31

源图片  $t_i$  之间的相似度  $S(q, t_i)$  计算公式为

$$S(q, t_i) = w_1 S_L(q, t_i) + w_2 S_E(q, t_i) \quad (6)$$

其中,  $w_1$  和  $w_2$  分别为底层特征相似度  $S_L$  和 Exif 信息相似度  $S_E$  的相似度权重(下角标  $L$  代表底层特征,  $E$  代表 Exif 信息)。  $w_1$  和  $w_2$  满足  $w_1 + w_2 = 1$ , 通过调整这两个权重的值就可以得到不同的检索结果。本文设  $w_1 = 0.7$ ,  $w_2 = 0.3$ , 获得了相对较好的检索结果。

## 3 实验结果以及分析

本实验利用的数据为取自一校园景色的包含如下 6 个概念的数据库:汽车、园林、草地、一高楼、一雕塑和一红色楼房。对于每个概念包含 50 张图片,总共 300 张图片。通过如下 4 组对比实验来说明本文所提出的算法的有效,即(1)基于全局底层特征的搜索;(2)融合了 Exif 信息的全局底层特征搜索;(3)基于区域的底层特征搜索;(4)融合了 Exif 信息的区域底层特征搜索。图 2 为实验演示系统,在该系统中,用户通过上传图片之后,就可以选择上述 4 种查询方式中的任何一个进行查询,系统则根据用户选择的查询方式,依据相似度由高到低返回用户的搜索结果。

本文采用搜索精度来评价算法的有效性。实验的结果显示,对于 6 个概念的检索本文算法的性能都优于传统的检索方法。图 3 和图 4 显示了汽车和雕塑两个概念的搜索精度。

从图 3、图 4 中可以看到,当搜索图片增大时,本文的算法获得了最好的搜索效果。对于汽车这一概念,本文算法比传统的方法搜索精度提高了 67.8%。对于雕塑这一搜索概念,当搜索数目较小时,本文的算法性能低于传统的基于内容的搜索方



图2 实验演示系统

Fig.2 The demo system of experiment

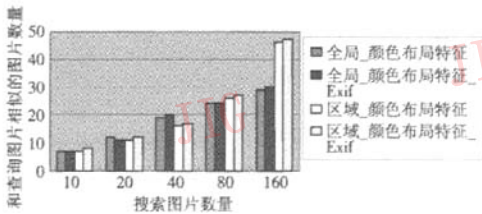


图3 汽车概念的搜索精度比较

Fig.3 Car concept image retrieval precision results

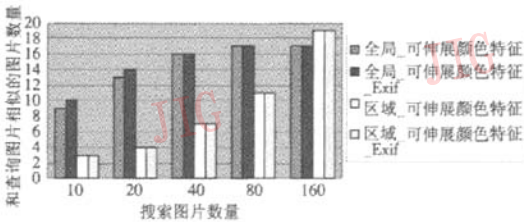


图4 雕塑概念的搜索精度比较

Fig.4 Sculpture concept image retrieval precision results comparison

法。通过分析雕塑图片发现,该语义图片中由于包括了很大的天空和绿化区域以及在建筑物雕塑的颜色上和高楼的类似,所以影响了系统的搜索性能。但是又发现随着搜索图片数量的增大,本文算法性能最终还是超过了传统的方法,因为本文算法融入了Exif信息,这也说明了整合Exif信息对于提高搜索性能是有效的。

## 4 结论

本文针对手机拍摄图片的日益增多的情况,设计了一个手机图片管理和搜索系统框架。基于此框架,本文提出了一种融合手机相机元信息的基于区域底层特征的图片搜索算法。与传统的基于内容的

全局图像搜索方法相比较,基于区域的方法能够在一定程度上解决语义鸿沟问题,同时本文算法利用图片的Exif信息来对所搜索的结果进行后处理。实验结果表明,本文的算法的性能优于传统的方法。后续工作将集中在用户的注意力选择方面,以便进一步改善搜索性能以及进一步研究Exif信息和图片的语义之间的关系方面。

## 参考文献 (References)

- 1 Davis M, Canny J, Van H N, *et al.* Mmm2: Mobile media metadata for photo sharing [A]. In: Proceedings of the 13th annual ACM international conference on Multimedia [C], Hilton, Singapore, 2005: 267 ~ 268.
- 2 Informedia System [EB/OL]. <http://www.informedia.cs.cmu.edu/>.
- 3 QBIC System [EB/OL]. <http://www.qbic.almaden.ibm.com/>.
- 4 Liu Y, Zhang D, Lu G, *et al.* Region-based image retrieval with high-level semantic color names [A]. In: Proceedings of the 11th International Multimedia Modeling Conference [C], Melbourne, Australia, 2005: 180 ~ 187.
- 5 Boutell M, Luo J. Bayesian fusion of camera metadata cues in semantic scene classification [A]. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], Washington, DC, USA, 2004: 623 ~ 630.
- 6 Luo J, Boutell M, Brown C. Pictures are not taken in a vacuum [J]. IEEE Transactions on Signal Processing Magazine, 2006, 23 (2): 101 ~ 114.
- 7 EXIF Standard [EB/OL]. <http://www.exif.org/>.
- 8 Deng Y, Manjunath B S. Unsupervised segmentation of color-texture regions in images and video [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23 (8): 800 ~ 810.
- 9 Shi J, Malik J. Normalized cuts and image segmentation [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97) [C], San Juan, Puerto Rico, 1997: 731 ~ 737.
- 10 Martinez J M. Overview of the MPEG-7 standard [R]. Technical Report 5.0, ISO/IEC, Singapore, 2001.
- 11 Rubner Y, Tomasi C, Guibas L J, *et al.* A metric for distributions with applications to image databases [A]. In: Proceedings of the 1998 IEEE International Conference on Computer Vision [C], Bombay, India, 1998: 59 ~ 66.