

中图分类号: TP301.6 文献标志码: A 文章编号: 1006-8961(2011)06-1045-06

论文索引信息: 冯振, 王宇新, 郭禾, 贾棋. 结合抠像技术的图像分类方法 [J]. 中国图象图形学报, 2011, 16(6): 1045-1050

结合抠像技术的图像分类方法

冯振^{1),2)}, 王宇新¹⁾, 郭禾²⁾, 贾棋²⁾

¹⁾(大连理工大学计算机科学与技术学院, 大连 116024) ²⁾(大连理工大学软件学院, 大连 116024)

摘要: 提出一种新的对象分类方法, 该方法将抠像技术应用到图像检索领域。给前景层图像块较高的权值, 背景层图像块较低的权值, 并对图像块按照权值大小排序, 图像间内容相似度比较时, 首先计算对应权值的图像块间的距离, 再将所有块间距离加权累加作为图像间距离。新方法的有效性在牛津大学花卉图像集上得到验证, 实验结果也表明, 相对于传统的计算对应位置的图像块间距离的方法, 新方法的检索准确度有明显的优势。

关键词: 图像检索; 抠像; 空间信息; 图像块

Objection classification via matting technique

Feng Zhen^{1),2)}, Wang Yuxin¹⁾, Guo He²⁾, Jia Qi²⁾

¹⁾(School of Computer Science and Technology, Dalian University of Technology, Dalian 116024 China)

²⁾(Software School, Dalian University of Technology, Dalian 116024 China)

Abstract: This paper presents a novel method for recognizing object categories. In this method, matting technique is applied to estimate foreground distribution information, which provides weights for image blocks. Blocks describing foreground are given higher weights, while blocks describing background are given lower weights, then all the blocks are sorted according to the assigned weight. We also conducted imagery comparison study. First of all, distances between blocks with corresponding weights are evaluated, and then the similarity between images can be obtained through adding all the distances between blocks by weights. Experiments on challenging object categorization image sets demonstrate the effectiveness of this method.

Keywords: image retrieval; matting; spatial cues; image blocks

0 引言

图像检索中的对象识别包括分类和检测两项任务, 分类用于判断一幅图像是否包含某类对象, 检测则要求标出这些对象的位置和大小。对象识别是理解图像和场景的关键, 具有广泛地应用前景, 可用于知识产权保护、新一代网上搜索、数字博物馆、医学和遥感图像的分析、犯罪与安全预防等诸多领域^[1]。

提出一种新的对象分类方法, 判定一幅照片中

是否包含感兴趣的特定对象。在这种分类任务中, 多种基于颜色、纹理和形状的图像特征描述已经被设计出来, 同时, 研究人员提出了多种方法试图将空间信息添加到图像特征表示中, 其大致可分为两类, 一类是基于图像空间固定划分的方法, 另一类是基于图像内容相关性聚类的方法。

在图像固定划分方法中, 具有代表性的如李国杰院士、何清法提出的基于分块主色的图像检索方法^[2], 该方法在检索时首先逐一计算对应位置分块的主色之间的距离, 然后将所有分块的距离进行某种程度的加权累加, 作为两幅图像之间的实际距离。

收稿日期: 2010-02-08; 修回日期: 2010-06-28

第一作者简介: 冯振(1987—), 男, 大连理工大学计算机科学与技术学院博士研究生, 主要研究方向为基于内容的图像检索与计算机视觉。E-mail: fengz@mail.dlut.edu.cn。

在刘洁敏、杨小康等人提出的基于局部颜色-空间特征的图像分类方法中,也采用了固定分块的策略提取图像颜色分布的空间信息^[3]。

基于图像内容相关性聚类的方法,其出发点在于图像可以划分为多个具有不同语义特征的区域。如王涛、胡事民和孙家广院士提出的基于颜色-空间特征的图像检索方法^[4],该方法将图像内容看成由若干对象组成的集合,首先利用图像分割得到主要对象,然后根据对象的颜色、位置和形状特征计算图像间内容的相似度,再进行检索。2009年,清华大学陈韬、胡事民等人研发出一种基于互联网的图像检索和合成的新算法 Sketch2Photo^[5],该算法在检索图像时,同样对图像进行了自动前景分割得到图像的空间信息。

对各种算法进行分析之后,我们认为,首先,图像的前景层和背景层都有助于图像最终检索准确度的提升;其次,图像的前景层信息对检索性能的影响会更大一些,因此应给与更多的关注度;第三,背景层信息不能被完全忽略,例如鸭子最有可能在湖水里游泳,而不是在屋顶上。

基于以上 3 个判断,我们认为有必要将图像前景层和背景层区域分离开来,不同于之前的图像分割方法,本文将抠像技术引入到图像检索领域中,计算图像中的每一个像素点的不透明度,并量化到 [0,1] 区间。像素点的不透明度越高,表示这个像素点越有可能描述的是图像的前景层对象,如图 1 所示,深红色表示该像素点最有可能描述的是前景层对象,深蓝色表示该像素点最有可能描述的是背景层环境。



图 1 前景层分布

Fig. 1 Foreground distribution

本文方法的贡献在于,首先,以往的固定划分方法在比较图像间内容相似度时,采取对应位置的图像块间的比较^[2-3],而本文方法采取的是对应权值的图像块间的比较;其次,为了计算图像块的权值,我们提取了前景层信息,不同于以往利用图像分割

提取前景层的方法,本文将抠像技术引入到图像检索领域中。

1 基于抠像技术的对象分类方法

综合已有的两类空间信息添加方法,我们提出结合抠像技术的图像分类方法:将图像固定划分,然后利用抠像技术计算每个图像块的权值(即计算图像块中所有像素点不透明度的平均值,并将其归一化),图像块的权值越高,其表示前景层的概率越高,对图像中所有图像块按照权值排序,并计算每个图像块的特征描述子。在比较图像间内容相似度时,首先逐一计算按权值排序之后的对应分块之间的距离,然后将所有分块间的距离加权累加,作为两幅图像之间的实际距离。本文方法的流程如图 2 所示。

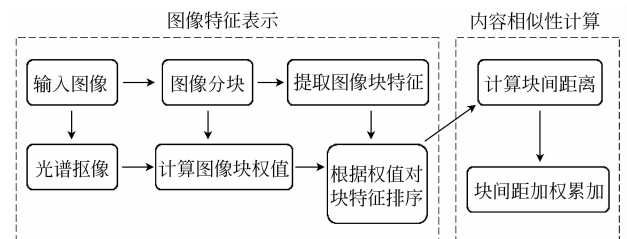


图 2 本文方法基本流程

Fig. 2 Flowchart for proposed method

1.1 图像空间信息提取

1.1.1 光谱抠像

数字抠像技术,是指从图像中分离出前景层,并对每个像素点求不透明度的过程。不透明度越高,代表该像素点描述的是前景对象的可能性越大,此前,抠像技术已经广泛应用于图像编辑^[6]、视频特效制作等领域。

抠像技术假设图像中每一个像素点 I_i 都是前景层颜色 F_i 和背景层颜色 B_i 的线性组合,组合系数 α_i 定义为不透明度:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i \quad (1)$$

若图像以 RGB 3 颜色通道表示,则式(1)中有 3 个已知量,7 个未知量,因此这是一个欠约束问题,如何减少人工参与,自动完成图像的抠像过程,引起了人们的广泛兴趣。采用光谱抠像技术^[7],据该技术的设计者称,光谱抠像是第一个不需要人工参与仍可获得较为理想的抠像效果的技术,其自动抠像的过程如图 3 所示。

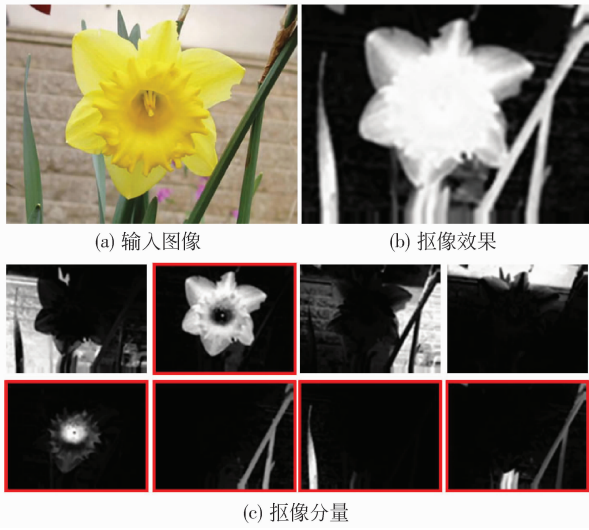


图3 光谱抠像

Fig. 3 Spectral matting

给定一幅输入图像(图3(a)),其每一个像素点都可以由 K 个图像层 F^1, F^2, \dots, F^K 的线性组合表示:

$$I_i = \sum_{k=1}^K \alpha_i^k F_i^k \quad (2)$$

抠像分量 α^k 为第 k 个图像层上所有像素点系数 α_i^k 的向量集合,其表示了第 k 个图像层对图像前景层的贡献率,如图3(c)所示。文献[7]证明,抠像分量可以通过拉普拉斯矩阵 L 的特征向量集合的线性组合求出, $L(i, j)$ 定义为

$$\sum_{q | (i, j) \in w_q} \left(\delta_{ij} - \frac{1}{|w_q|} \left(1 + (I_i - \mu_q)^T \left(\sum_q + \frac{\varepsilon}{|w_q|} I_3 \right)^{-1} (I_j - \mu_q) \right) \right) \quad (3)$$

式中, \sum_q 是 3×3 的协方差矩阵, μ_q 是 3×1 维的颜色均值向量, I_3 是 3×3 的单位矩阵, δ_{ij} 为 Kronecker 系数,而 $|w_q|$ 为 q 邻域窗口内的像素个数,关于式(3)更深入的分析,可参考文献[7]。

在求得所有的抠像分量 $\alpha^k, k=1, 2, \dots, K$ 后,假设其中 $\alpha^{k_1}, \alpha^{k_2}, \dots, \alpha^{k_n}$ 为前景层分量,则图像前景层可以表示为

$$\alpha = \alpha^{k_1} + \dots + \alpha^{k_n} \quad (4)$$

定义代价函数 $J(\alpha) = \alpha^T L \alpha$, 给定所有可能的前景层分量集合,计算代价函数最小的一类分量集合,即作为图像的最终前景层分量集合,其在图3(c)中由红框标注,由此便可以得到图3(b)的抠像效果。

光谱抠像面临的一个问题,是其很高的计算复杂度,建议在不改变原抠像算法的前提下,使用插分算法缩小输入图像的尺寸:给定一幅输入图像,输出的像素值,为该像素 4×4 邻域内所有像素值的加权平均值,插分算法迭代执行直到图像缩减到规定尺寸为止。本文方法并未降低抠像的计算复杂度,但是实验证实,随着输入数据量的降低,抠像的系统资源占用量取得了明显的改善。

1.1.2 图像块权值计算

使用光谱抠像技术得到图像每个像素点的不透明度,将图像划分为 N 个固定大小的图像块,计算每个图像块内像素点不透明度的平均值 w_1, w_2, \dots, w_N 作为各图像块的权值,定义权值中最大的值为 w_{\max} , 最小值为 w_{\min} , 对每一个权值 w_i 归一化:

$$w'_i = \frac{w_i - w_{\min}}{w_{\max} - w_{\min}} \quad (5)$$

1.2 图像块特征提取

本文实现了两种图像特征描述子,即颜色矩^[8]和致密 SIFT^[9]。

在颜色矩特征描述子的实现中,由于实验图片都是在户外自然光环境下拍摄得到的,光照条件随着天气和太阳照射角度改变而改变,为了降低光照的影响,我们将图像由 RGB 颜色空间转换到 YCbCr 颜色空间。接着按照 1.1.2 节分块方法将图像划分为 N 个子块,在每一个图像块上计算其在 3 个颜色通道的一阶矩(mean)、二阶矩(variance)和三阶矩(skewness)。由此得到 N 个图像块,每个图像块用 9 维颜色矩向量表示。

不同于传统的先检测局部极值点再对极值点提取 SIFT 特征的方法^[10](亦称之为稀疏 SIFT),采用致密 SIFT 图像特征^[9]。在图像上,等间距地选择样本点,对所有样本点提取 SIFT 特征,根据特征袋算法^[11],使用 K 均值聚类^[12]将所有 SIFT 特征聚类到 200 个聚类中心。按照 1.1.2 节分块方法将图像划分,统计每一个子块上样本点的 SIFT 特征,由此得到 N 个图像块,每个图像块用 200 维的 SIFT 直方图表示。

1.3 图像间内容相似度计算

输入一幅图像,根据图像块权值由大到小对图像块重排序,图像特征便可以表示为图像块特征与其对应权值的组合,如图4所示。

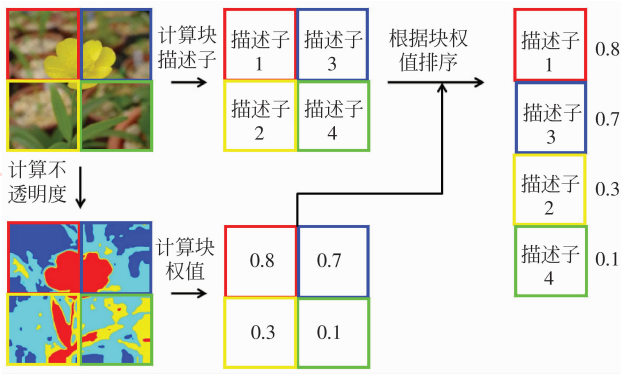


图 4 图像特征表示方法

Fig. 4 Image feature representation

给定两幅图像 p 和 q , 其根据权值排序之后图像特征表示分别为 $\langle (f_{p1}, w_{p1}), (f_{p2}, w_{p2}), \dots, (f_{pN}, w_{pN}) \rangle$ 和 $\langle (f_{q1}, w_{q1}), (f_{q2}, w_{q2}), \dots, (f_{qN}, w_{qN}) \rangle$, 其中 f_{pi} 和 f_{qi} 分别表示为图像 p 和 q 第 i 个图像块的特征描述子, w_{pi} 和 w_{qi} 分别表示为第 i 个图像块的权值, 且有 $w_{p1} \geq w_{p2} \geq \dots \geq w_{pN}, w_{q1} \geq w_{q2} \geq \dots \geq w_{qN}$, 即图像块特征描述子是按其权值大小排列的。定义 d_{pq}^i 为图像 p 和 q 对应第 i 块特征描述子的欧拉距离, 则图像间内容相似性可表示为

$$D_{pq} = \frac{(w_{p1} + w_{q1})}{2} d_{pq}^1 + \frac{(w_{p2} + w_{q2})}{2} d_{pq}^2 + \frac{(w_{p3} + w_{q3})}{2} d_{pq}^3 + \dots + \frac{(w_{pN} + w_{qN})}{2} d_{pq}^N \quad (6)$$

由式(6)可知, 前景层区域提取出的图像块特征描述子的权值较大, 因此在最终的整幅图像的描述子中, 前景图像信息得到更多地关注, 而背景信息又没有完全被忽略。

2 实验和数据分析

实验的硬件平台为 Intel Core i5-750 CPU, 4 内核, 主频 2.67 GHz; 3 G 内存。软件平台为 Windows XP SP2 上 MATLAB 7.6.0.324 (R2008a), MATLAB 自带的 JAVA 虚拟机版本为 1.6.0。

相关实验采用英国牛津大学花卉图像数据库^[13], 并在其中选取 6 类共计 360 幅照片作为实验图片(如图 5), 数据库中所有照片皆拍摄于户外自然光环境下, 每张图片的拍摄角度不固定, 前景层多个对象间存在重叠遮挡现象, 这使得对象分类任务更富挑战性。



图 5 花卉图像

Fig. 5 Images from the flower category dataset

实验共分为 3 部分: 首先通过举例和查全率-查准率曲线图的方式证实本文提出的图像分类方法的优越性; 然后给出图像分类结果的混淆矩阵, 最后是光谱抠像在改进前后系统资源消耗的比较。

我们称文献[2-3]采用的对应位置图像块比较方法为改进前方法, 本文提出的对应权值图像块比较方法为改进后方法, 则图 6 比较了改进前方法与改进后方法计算出的 SIFT 特征描述子。

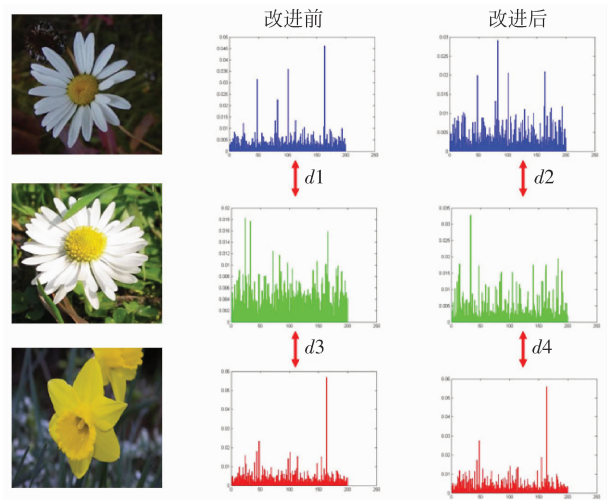


图 6 两种特征描述子比较

Fig. 6 Descriptors comparison

欧拉距离 $d1 = 464.5858$, $d2 = 144.1799$, $d3 = 939.4708$, $d4 = 468.0593$, 在改进前类间类内距离比为 $d3/d1 = 2.0222$, 而在本文方法中, 类间类内距离比增加到 $d4/d2 = 3.2464$ 。这就在另一个角度验证了本文方法能够获得类间区分度更高的特征描述子集合。

查全率 (recall) 和查准率 (precision) 是信息检索中的标准评价方法, 现在已经越来越多地使用在

图像检索当中,其定义如下。

$$\begin{aligned} \text{查全率} &= \frac{\text{有关联的正确检索结果}}{\text{所有有关联的结果}} \\ \text{查准率} &= \frac{\text{有关联的正确检索结果}}{\text{所有检索到的结果}} \end{aligned} \quad (7)$$

对任意一幅输入图像,计算其与数据集中所有图像的相似度,由此可以按照相似度对所有图像进行排序,在图像序列中,当取值指定数目的正确检索结果时,便有对应的查全率与查准率,由此随着指定正确检索图片数目的改变,便可以得到查全率-查准率变化曲线。

实验中,对每一类随机选择 5 幅图片,共计 30 幅图片作为输入图片,计算每一幅图片的查全率-查准率曲线,从而得到平均查全率-查准率曲线,图 7 为图像分块数 $N = 16$,块特征取颜色矩时的曲线图;图 8 为图像分块数 $N = 4$,块特征取致密 SIFT 时的曲线图。同前文表述,称文献[2-3]采用的对应位置图像块比较方法为改进前方法,本文提出的对应权值图像块比较方法为改进后方法,可以看出在相同的查全率下,本文方法得到的查准率占有比较明显的优势。

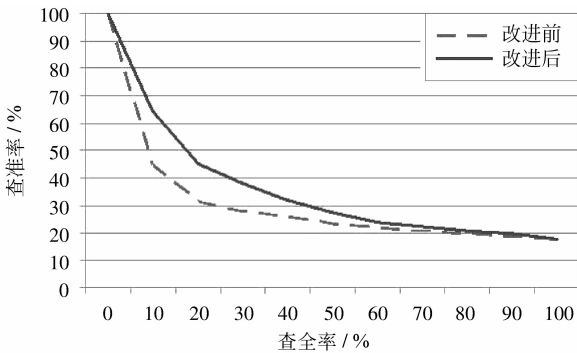


图 7 颜色矩下查全率-查准率曲线

Fig. 7 Recall-precision curve for color moments

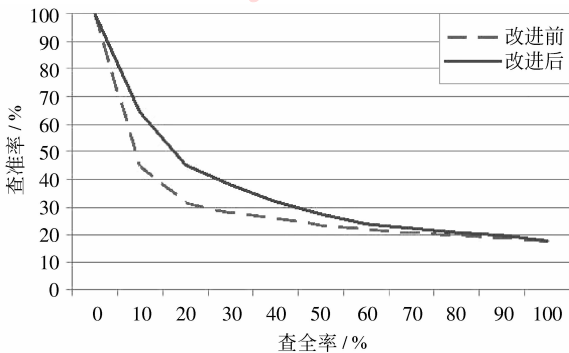


图 8 致密 SIFT 下查全率-查准率曲线

Fig. 8 Recall-precision curve for dense SIFT

混淆矩阵 (confusion matrix) 也是图像分类中常用的标准方法之一,其量化描述了一类图像被预测为某类图像的百分比。定义图像集中有 m 类图像, n_{ij} 为第 i 类图像中被预测为第 j 类的百分比,若 $i = j$ 则 n_{ij} 表示第 i 类被正确预测的百分比,于是有混淆矩阵

$$CM = \begin{Bmatrix} n_{11} & \cdots & n_{1m} \\ \cdots & n_{ij} & \cdots \\ n_{m1} & \cdots & n_{mm} \end{Bmatrix} \quad (8)$$

实验中,对每一幅图像计算其与其他 359 幅图像的相似度,使用 K 近邻算法^[14]预测每一幅图像类别,由此统计得到本文方法的混淆矩阵。图 9 为颜色矩下的混淆矩阵,图 10 为致密 SIFT 下的混淆矩阵,图中颜色越浅代表其数值越高。

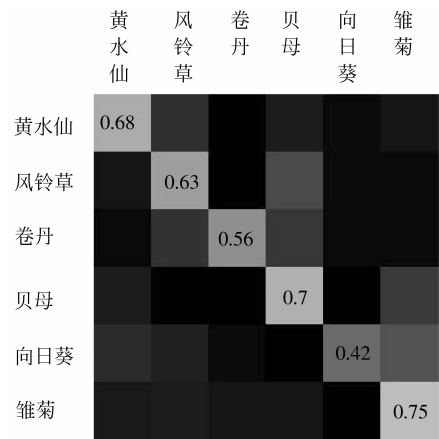


图 9 颜色矩下混淆矩阵

Fig. 9 Confusion matrix for color moments

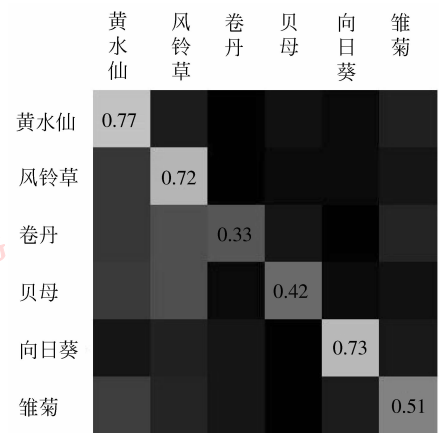


图 10 致密 SIFT 下混淆矩阵

Fig. 10 Confusion matrix for dense SIFT

在对一幅典型的输入图像执行缩略前后,完成光谱抠像所需要的系统资源消耗量的比较(如表 1 所示),可以看出本文使用的图像迭代缩略策略对抠像技术性能的提升作用明显。

表 1 光谱抠像资源消耗比较

Tab.1 System resource consumption for spectral matting

	缩略前	缩略后
图像尺寸	550 × 660 × 3 像素	88 × 114 × 3 像素
图像大小	7.6 MB	235 KB
处理时间	内存溢出,运行终止	9.7 s
内存占用	1.26 GB	675 MB

缩略前的图像在处理过程中,会出现内存溢出而运行终止,我们分析认为,该异常是由 Matlab 自带的 Java 虚拟机内存过小(只有 128 MB)所致。

3 结 论

提出一种新的对象分类方法,描述如何将抠像技术应用到图像检索领域中。本文方法考虑了不同图像区域(前景区域、背景区域)对图像检索性能的影响,并将其作为空间信息以权值的形式加入到图像特征表示中。未来的工作将会更多的关注于设计出某种更准确的自动抠像技术,另一个方向就是选择抠像之外的其他方法来提取出图像中关注度设定更高的区域。

参考文献 (References)

- [1] Zhou Quanming, Geng Guohua, Wei Na. Content-Based Image Retrieval [M]. Beijing: Qinghua University Press, 2007: 6-7. [周明全, 耿国华, 韦娜. 基于内容图像检索技术 [M]. 北京: 清华大学出版社, 2007: 6-7.]
- [2] He Qingfa, Li Guojie. An approach to combining dominant color of partition with relevance feedback in image retrieval [J]. Journal of Computer-aided Design & Computer Graphics, 2001, 13(10): 912-917. [何清法, 李国杰. 综合分块主色和相关反馈技术的图像检索方法 [J]. 计算机辅助设计与图形学学报, 2001, 13(10): 912-917.]
- [3] Liu Jiemin, Yao Yu, Zhang Rui, et al. Image semantic concept detection based on local color-spatial feature [J]. Journal of Image and Graphics, 2008, 13(10): 1890-1893. [刘杰敏, 姚豫, 张瑞, 等. 基于局部颜色-空间特征的图像语义概念检测 [J]. 中国图象图形学报, 2008, 13(10): 1890-1893.]
- [4] Wang Tao, Hu Shimin, Sun Jianguang. Image retrieval based on color-spatial feature [J]. Journal of Software, 2002, 13(10): 2030-2036. [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): 2030-2036.]
- [5] Chen Tao, Cheng Mingming, Tan Ping, et al. Sketch2photo: internet image montage [C]// ACM SIGGRAPH ASIA 2009. Yokohama, Japan: Association for Computing Machinery, 2009, 124: 1-10.
- [6] Fu Xinyuan, Guo He, Wang Yuxin, et al. Seamless cloning algorithm based on image matting technique [J]. Journal of Image and Graphics, 2008, 13(6): 1082-1089. [傅新元, 郭禾, 王宇新, 等. 基于抠像技术的图像无缝融合算法 [J]. 中国图象图形学报, 2008, 13(6): 1082-1089.]
- [7] Levin A, Rav-Acha A, Lischinski D. Spectral matting [J]. IEEE Pattern Analysis and Machine Intelligence, 2008, 30: 1699-1712.
- [8] Stricker M, Orengo M. Similarity of color images [C]// Proceedings of SPIE Storage and Retrieval for Still Image and Video Databases. San Diego, CA, USA: SPIE, 1995, 2420: 381-392.
- [9] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, United states: IEEE Computer Society, 2005: 886-893.
- [10] Lowe D. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [11] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]// Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, United states: IEEE Computer Society, 2006: 2169-2178.
- [12] Macqueen J B. Some methods for classification and analysis of multivariate observation [C]// Proceedings of the Fifth Symposium on Math, Statistics, and Probability. Berkeley, CA, United States: University of California Press, 1963: 281-297.
- [13] Nilsback M E, Zisserman A. A visual vocabulary for flower classification [C]// Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, United states: IEEE Computer Society, 2006: 1447-1454.
- [14] Cover T M, Hart P E. Nearest neighbor pattern classification [J]. IEEE Translation Information Theory, 1967, 13(1): 21-27.