

中图法分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)04-0503-07

论文索引信息: 贺广南, 杨育彬, 阮佳彬, 林金杰. 基于视觉一致性的图像检索 [J]. 中国图象图形学报, 2011, 16(4): 503-509

基于视觉一致性的图像检索

贺广南, 杨育彬, 阮佳彬, 林金杰

(南京大学软件新技术国家重点实验室, 南京 210093)

摘要: 提出一种新的彩色图像分割方法, 将图像分割成具有明显视觉一致性的区域, 这种一致性能够模拟人观察图像时的视觉感受, 例如图像中的一片区域具有相同的颜色、纹理。对这样的一致性区域建立特征描述符, 如颜色编码、连通系数、面积比例, 其中颜色编码是通过将像素在 HSI 颜色空间中量化得到, 进而为整幅图像建立特征描述; 然后将这种特征描述用于图像的检索。实验结果表明, 这种方法不仅能够很好地模拟图像所带给人的视觉感受, 而且对具有视觉一致性的图像检索效果也有很好的表现。

关键词: 视觉一致性区域; 图像分割; 连通系数; 图像检索

Image retrieval based on visual consistency

He Guangnan, Yang Yubin, Ruan Jiabin, Lin Jinjie

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093 China)

Abstract: In this paper, firstly a new method for color image segmentation is proposed. The segmentation algorithm can segment an image into several regions with high visual consistency. The segmented results can simulate the feelings of human observing images, like an image have several regions with same colors or texture. After obtaining consistent regions, some image region properties like color coding, connectivity coefficient, and area ratio are then extracted as descriptors which can be used for content-based image retrieval. The color coding was obtained by quantification the pixels in the HSI color space. The experimental results shows that the proposed method can not only capture the visual impression of an image, but also have a good performance in image retrieval.

Keywords: visual consistency regions; image segmented; connectivity coefficient; image retrieval

0 引言

一直以来图像检索是一个非常活跃的研究领域, 大量的方法都是基于图像的底层视觉特征, 例如提取图像中的颜色、纹理、形状等特征, 然后利用这些视觉特征来做图像的相似性匹配, 从而产生许多以图像内容为基础的图像检索的相关技术。图像检索与图像类别的分类是密不可分的, 近年来很多研究者, 努力研究能够识别不同种类的图像的方法, 出

现了各种各样的模型。例如, Chris 等人^[1]将图像看成一个“词袋”, 每一个“词”表征图像的某一个特征, 这样就可以和文字信息检索很好地结合起来, 最终利用机器学习的方法如贝叶斯分类和支持向量机等, Lowe^[2]通过提取图像中局部的尺度不变性的特征, 用于物体的识别, 提取出的特征统称为 SIFT 向量, 每一个 SIFT 特征对应于图像中的一个 SIFT 关键点, 实验表明用 SIFT 向量来描述图像特征, 并用于图像的匹配、检索, 具有很好的稳定性^[3]。Li 等人^[4-5]中将图像的场景划分为一个个的局部区域,

收稿日期: 2009-06-19; 修回日期: 2009-12-08

基金项目: 国家重点基础研究计划(973)项目(2010CB327903); 国家自然科学基金项目(60875011, 60975043, 60723003, 61021062); 江苏省自然科学基金创新人才(学术带头人)基金资助项目(BK2010054)。

第一作者简介: 贺广南(1985—), 男, 南京大学计算机科学与技术系 2008 级硕士研究生, 研究方向为图像检索。E-mail: gnhe2009@gmail.com。

每个局部区域表征一个关键词,一类图像中的关键词就表征了该类的主题,如室内、室外,通过统计学习的方法建立产生式的模型,用于图像的分类和场景的识别。Yang 等人^[6]中提出一种基于连通颜色区域的图像检索方法,该方法能够有效结合图像的纹理信息及其颜色构成和分布信息,具有良好的检索效果和性能。

上述的各种方法都没有考虑人自身对图像的视觉感受,只是根据大量的图像特征,建立起一定的辨别模型,利用机器学习的方法来完成图像的分类或检索;另外图像的分割应是图像检索过程中一个关键的步骤,所以将图像分割成一个个有意义的区域,是图像检索面临的首要难题,大量的实验表明良好的图像分割能够极大地提高目标识别和检测的准确率^[7]。本文试图从人的视觉特性出发,将图像进行自动的视觉一致性区域分割,然后建立图像中一致性区域的描述,最终将这样的描述应用于图像检索中去,其实这样的一种分割更好地表征了图像区域特征,获取了图像全局的拓扑结构;另外通过这样的分割也为基于语义的图像检索提供了很好的方法。

1 图像颜色空间的量化编码

HSI 模型^[8]最早是由 Munseu 提出的,对于人类来说它们看起来更加自然和直观,反映了人的视觉系统观察彩色的方式,在艺术创作上经常使用 HSI 模型。HSI 模型中, H 表示色调 (Hue), S 表示饱和度 (Saturation), I 表示亮度 (Intensity, 对应像素亮度),也可将 HSI 称做 HSL (Lightness) 或 HSB (Brightness)。本文运用这个颜色模型基于以下两个重要的事实,1) HSI 颜色模型中 I 分量与图像的彩色信息是相互独立的;2) HSI 颜色模型中 H 和 S 分量与人感受颜色的方式是紧密相连的。因此,使得 HSI 模型非常适合开发基于彩色描述的图像处理算法,也能很好地模拟人的视觉系统感知彩色特性。由 RGB 彩色空间转换到 HSI 彩色空间的细节请参考文献^[8]。

要获得一致性的区域,需要从新量化原图像的颜色空间,原图像的颜色空间都是基于 RGB 颜色模型的,每个像素点包括 R 、 G 、 B 3 个颜色通道,将图像 RGB 颜色模型转换到 HSI 颜色模型中,然后做进一步的处理。量化在 HSI 彩色空间上进行,方法是将 HSI 彩色空间非等间距量化到 32 维彩色空间中,

相当于一个颜色直方图。每一个数值代表了一种颜色,对于视觉系统而言,不同数值对于人类来说视觉感受是不一样的。最终将 HSI 空间降至 32 维的颜色空间,也相当于从新对图像进行了编码。编码的过程如下:

- 1) if $I \leq 0.2$, 认为颜色是黑色的, $code = 0$ 。
- 2) if $S \leq 0.1$ & $I > 0.2$, 这样的颜色将会按照亮度 I 划分为 3 种颜色,分别为深灰色 (0.2, 0.5]、浅灰色 (0.5, 0.8] 和白色 (0.8, 1]。 $code = 1, 2, 3$ 。
- 3) 其余的颜色划分为彩色,共有 28 种彩色。

$$h' = \begin{cases} 0 & H \in (330, 360] \cup [0, 20] \\ 1 & H \in (20, 45] \\ 2 & H \in (45, 75] \\ 3 & H \in (75, 165] \\ 4 & H \in (165, 200] \\ 5 & H \in (200, 270] \\ 6 & H \in (270, 330] \end{cases}$$

$$s' = \begin{cases} 0 & S \in (0.1, 0.45] \\ 1 & S \in (0.45, 1] \end{cases}$$

$$i' = \begin{cases} 0 & I \in (0.2, 0.5] \\ 1 & I \in (0.5, 1] \end{cases}$$

$$code = 4 + 4h' + 2s' + i'$$

其中 H, S, I 分别代表所在像素点的在 HSI 颜色空间的数值。重新对图像编码后的效果如图 1 所示。

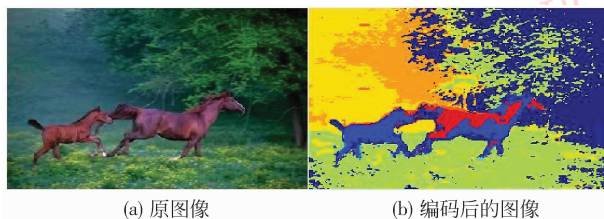


图 1 原图像和编码后的图像

Fig. 1 Original image and the encoded image

重新编码后的图像与原图像相比,数据量已大大减小,而且数值的变化范围也缩小到了 $[0, 31]$ 。图 1 中。重新编码后的图像矩阵中的数据变化范围和维数的降低,这都为下一步的处理算法提供了有力保证。

2 视觉一致性区域定义及提取算法

视觉一致性区域,指将图像重新编码量化到 32 维的数据空间中,在这样的空间中寻找最可能大的

连通区域,这样的连续性区域就代表了图像中的一个视觉一致性区域。区域的连通一致性用定义的连通系数来度量,例如图像的一棵树、一片草地等,本身它的连通性系数可能不大,但是这样的区域看起来却是一致的,即一片树叶或一片草地等。对于连通性系数较高的区域可能就是一个具体的事物,因为它表面的色彩变化不大,这样的区域具有高度的连通性、一致性。重新编码过的图像设为 I , 我们通过将区域的连通性分等级的方法来定义图像的连通系数, I 中每一个码值的连通性分为 8 个等级,如图 2 显示了所有的连通等级,以黑色点为中心,形成一个 3×3 的窗口,所有黑色点代表同一个码值,未标注的点表示其他的码值。其中 I 中 $code = i$ 的连通性系数可通过下式计算:

$$\alpha_i = \frac{\sum_{k=1}^8 c_i^k}{8 \times \sum c_i} \quad i \in [0,31]$$

k 表示连通的等级,如一连通、二连通等, c_i 表示 $code = i$; c_i^k 表示在 k 连通性等级下 $code = i$ 的像素个数; $\sum c_i$ 表示 $code = i$ 的所有像素个数,因为连通系数的计算是逐点进行的,因此并不需要考虑周围点的排列情况。

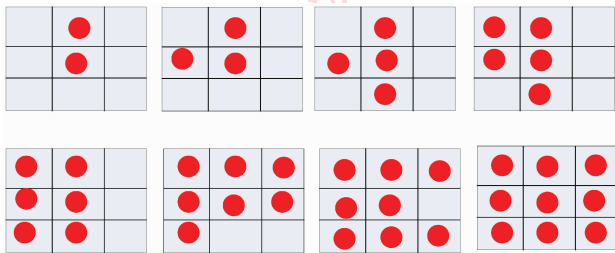


图 2 点的 8 种连通等级

Fig. 2 The eight connectivity level of point

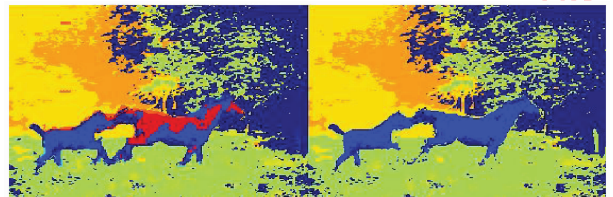
对于图 2 中的模板连通等级 1 ~ 8 的连通性系数分别为 0.125, 0.25, 0.3125, 0.4, 0.4583, 0.4643, 0.4688, 0.5556。

显然当连通等级增大时连通性系数单调上升,因为图 2 模板中除了第 8 个的中心点外,其他的点都被看成了边缘点,所以它们的连通性系数都较低,若将它们都看成非边缘点,则它们的连通性系数分别为 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1。

连通系数等于 1 是一种极限的情况,因为如果

当某一码值区域的点都为非边缘点时(此时也包含了中心点),也只有在该码值区域的面积无限大的情况在下才会发生,也即 $\lim_{r \rightarrow \infty} \alpha_i = \frac{8 \times \sum c_i}{8 \times \sum c_i} = 1$, 但这种情况在实际问题中是不会出现的,说明图像中所有码值的连通系数都将会小于 1。

实验中表发现, I 中会出现一些数量极小的码值,这些数量极小的编码值的出现,大多是因为图像中局部区域的明暗度的变化,或色调和饱和度的微小波动引起的,所涉及面积即该编码值的总个数比较少,所以需要对其 I 进行一步关键性的处理——非主一致性区域的抑制,使之融合到空间上相邻且面积大、连通系数较高的区域当中。设定阈值为 $\beta = 0.025$, 即当 $S_{c_i} < S \cdot \beta$ 时,则 $code = i$ 的像素就需要融合, S_{c_i} 表示 $code = i$ 的面积, S 表示图像的面积。即当区域面积小于图像总面积的 1/40 时就要融合掉该区域。由图 2 可知,当 I 中的点连通性不高时,其必与其他的点相邻接,非主一致性区域的抑制,就是要找出这些相邻的点,并比较连通性的大小以决定将其融合到哪个主一致性区域当中,找出的这些相邻的点满足: $S_{c_i} \geq S \cdot \beta$, 在找出的这些点中,可能会对不同的码值,选取与需要融合的 $code$ 连通性最好的主一致性区域的码值作为最终的融合值,融合效果如图 3 所示。



(a) 编码后的图像

(b) 融合后的图像

图 3 编码后的图像和融合后的图像

Fig. 3 The encoded image and the fused image

对重新编码后的图像融合完之后,如图 3 所示可以清楚地看到原图中的基于视觉的一致性区域,可参照图 1。这确实反映了图中的几个视觉一致性的区域,要描述这样的一致性区域,我们需要描述这样的区域的参数,求出 $code$ 值后,该 $code$ 值所代表的区域的连通系数和面积的大小都可以求出,图 3 中融合后的图像数据如表 1 所示;将表 1 中的码值,从图像 I 中提取出来,显示在二值图像中如图 4 所示。

表 1 一致性区域的参数

Tab. 1 The parameters of the consistent regions

码值 c_i	连通系数 α_i	面积 S_{c_i}
0	0.848 857 780 831 33	0.296 122 233 072 91
4	0.919 848 419 848 42	0.071 136 474 609 38
16	0.864 066 154 910 55	0.308 766 682 942 71
18	0.478 696 285 506 19	0.041 900 634 765 63
20	0.898 769 441 630 42	0.170 705 159 505 2
22	0.848 568 231 640 48	0.111 368 815 104 16

从图中可以看出,连通性系数较小的 $code$ 值具有明显的纹理特征,如 $code = 18$ 时对应了原图中的树叶、草地等, $code = 4$ 时,连通性系数最大,则对应了原图中明显的一致性区域——两匹马。融合后的图像只有 6 个一致性区域,在这 6 个一致性区域当

中有的区域的连通系数较小,但没有被融合掉,其原因是该区域所占源图像的面积较大,如图 4 中的 $code = 8$ 。从整体上这 6 个区域很好地描述了该图像的拓扑结构,也即空间上的分布情况,连通系数大的空间上分布比较紧密,小的空间上分布比较散乱。整个算法可描述为:

1) 根据第 2 部分介绍的量化方法,将原彩色图像非等间距量化,生成重新编码的图像 I 。

2) 对图像 I 中所有的 $c_i, i \in [0, 31]$; 如果满足 $S_{c_i} > 0$, 则根据式(1)计算 α_i 。

3) 对图像 I 中所有的 $c_i, i \in [0, 31]$; 如果满足 $S_{c_i} < S \cdot \beta$, 则将 c_i 融合掉,即非主一致性区域的抑制。

4) 输出处理后的图像 I 。

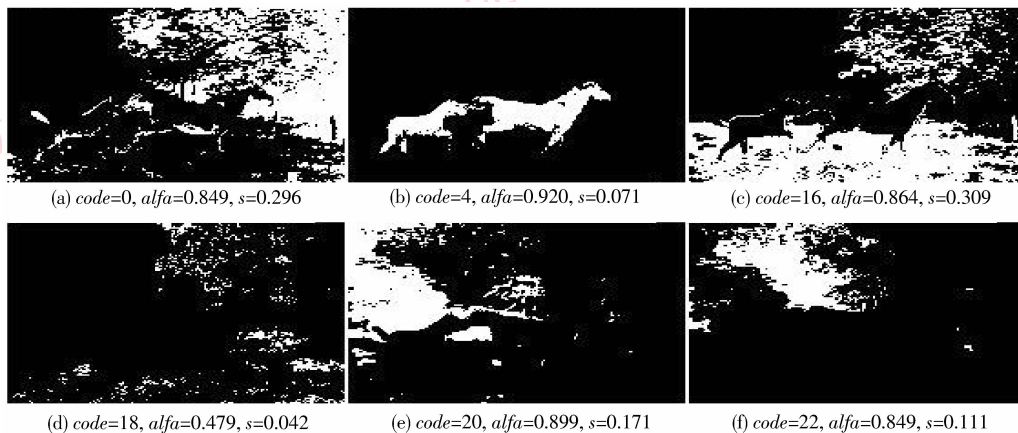


图 4 各个码值的二值图像

Fig. 4 Binary image of each code value

实验表明图像经过上述处理过程之后,图像中的一致性区域能够准确辨别,原图像已被分割成了几个一致性的区域,图像的分割融合效果如图 5 所示。

下节将要具体讨论怎样应用分割图像后所得的数据,用于图像检索当中。

3 实验分析

实验中所用到的图像集是 Corel image dataset^[9], 该图像库在文献[10]中作为实验图像库。对每一幅图像经过处理后,我们可以得到该图像的一致性数据描述,数据类型同表 1。我们应用不同的数据量在此数据库上做了两个检索实验。

实验 1 设任意两个图像 I_1, I_2 , 经过处理后, 它们的码值集合分别为 C_{I_1}, C_{I_2} , 定义集合交集 $C_{1,2} = C_{I_1} \cap C_{I_2}$, 若 $C_{1,2} = \emptyset$, 说明两幅图像在视觉感受上是完全不同的图像; 否则, $C_{1,2} \neq \emptyset$, 计算 $S_{inter} = \sum \min(s_i^{I_1}, s_i^{I_2}), i \in C_{1,2}, s_i^{I_1}, s_i^{I_2}$ 分别代表图像 I_1, I_2

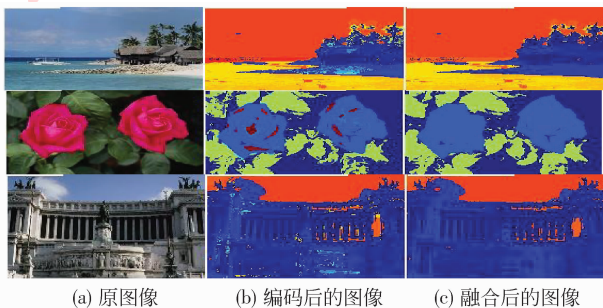


图 5 编码处理过的图像

Fig. 5 Processed images

在 $code = i$ 时的面积的百分比,在不考虑连通系数 α_i 时,将相似性的阈值 $T_{threshold} = 0.5$,即当 $S_{inter} > T_{threshold}$ 时,则该两幅图像具有视觉上的一致性,用 Corel image dataset 中的 1000 个图像集,图像集中共有 10

个类别,每个类别共有 100 张图片。基于以上的判别标准,计算出每一类的查全率(召回率)和查准率(精度)的平均水平,得到其 PR 直方图如图 6 所示,图 7 显示的是查询 horses 类所返回的结果。

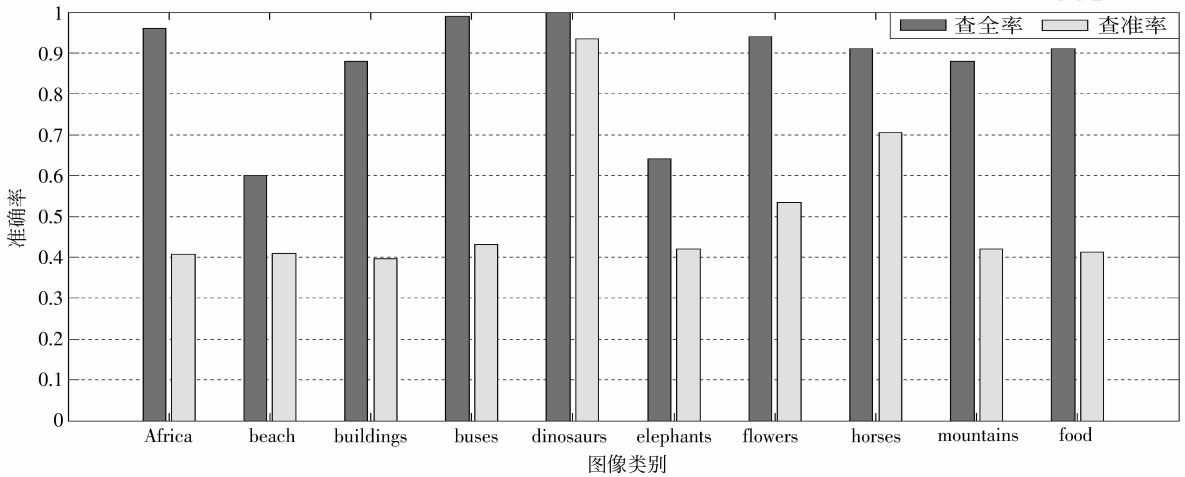


图 6 实验 1 所得 PR 直方图

Fig. 6 Histogram obtained from experiment one



图 7 查询结果图

Fig. 7 Results of the retrieval experiment

其中左上角第 1 幅图像为输入的查询图像,从查询结果可以看出,误查的图像中也都包含有与查询图像相一致的区域,如绿色的草地、树叶等,算法对该类图像的效果较好的原因也是显而易见的,首先在这类图像中,场景简单,一致性区域明显,且连

通系数较高,所以查询结果会好,其他几类中连通性系数较高的区域几乎没有,所以查询起来效果会差。产生这一现象的另一原因是查询中只用了两幅图像中相交码值的面积数,并未体现连通性系数的作用,毕竟连通性系数反映了某一一致性区域的分布情

况。导致查准率较低的原因是由图像库的分类问题,因不同类的图像却包含有相同的一致性区域,在图 7 中都有绿草地,这是由于待查图像中包含这样的一致性区域。

实验 2 上述实验中并未使用一致性区域的连通性系数,虽然两幅图像中共同一致性区域的面积大于了阈值 $T_{\text{threshold}}$,但它们的分布并不一定一样,就像前面讨论的,连通系数小的一致性区域可看成基于纹理的一致性区域,则它与非纹理的一致性区域是显然不同的。实验 2 中在实验 1 的基础上,引入连通系数的方差,即利用两个一致性区域的连通系数的均方差来判定两个一致性区域的分布情况,若

方差大于一定的阈值 $\sigma_{\text{threshold}}$,则无论共同区域的面积再大,它们也将是不同种类的图像,实验应根据图像种类的不同 $\sigma_{\text{threshold}}$ 也是变化的,实验过程中令 $0.1 \leq \sigma_{\text{threshold}} \leq 0.3$,例如针对区域分布较为散乱的图像类别“buildings”可将阈值放大到 0.3,而对一些区域分布较为集中的图像类别“flowers”可将阈值设为 0.1。实验证明阈值在这个区间中效果达到最优,并且对查询结果的影响不大,基本上保持在一个稳定的水平。查询结果的 PR 直方图如图 8 所示,从图 8 可以看出,各个类的查准率都有所上升,尤其是“flower”类,这也说明对于有明显连通性区域的图像,连通性系数对检索准确性的提高有很大帮助。

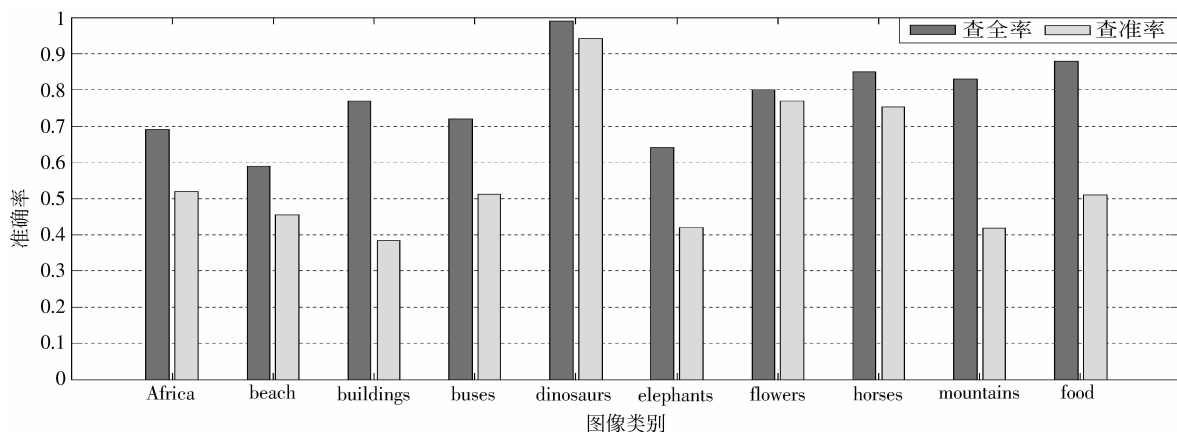


图 8 实验 2 所得 PR 直方图

Fig. 8 Histogram obtained from experiment two

由于本文使用的是文献[10]中的数据库,为了验证本方法的有效性,将本文所得的检索准确率结

果与文献[10]中的结果的对比如图 9 所示,该准确率结果是每个图像类别的 50 幅图像的检索平均值。

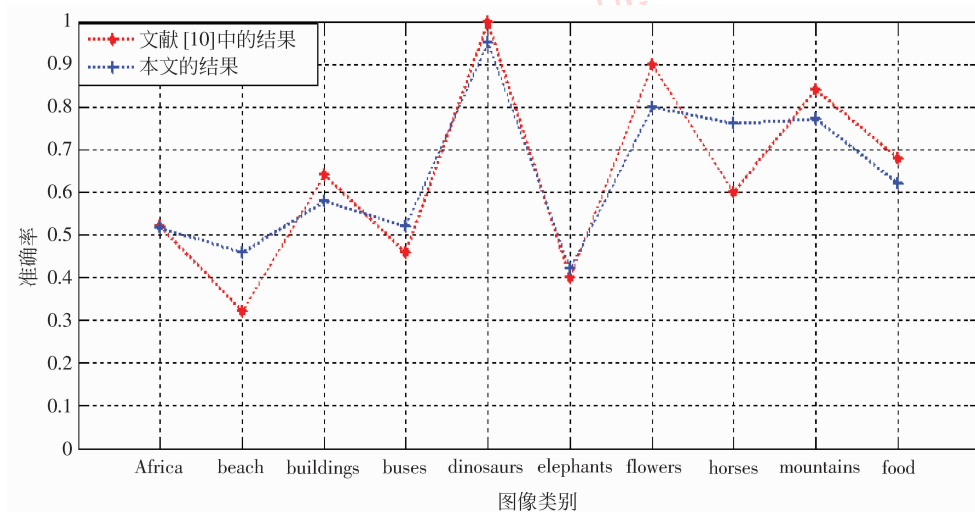


图 9 实验结果对比

Fig. 9 Results contrast for the two algorithms

从图9中可以看出两种方法在不同的图像类别上检索性能各有优劣,例如本文的方法在类别“beach”、“horses”、“buses”上实验结果好于文献[10]的方法;而文献[10]中的方法在类别“buildings”、“flowers”、“mountains”上好于本文所使用的方法;在类别“Africa”、“dinosaurs”、“elephants”上两者相差无几。但文献[10]中利用到概率统计模型^[11]来处理图像和建立图像的描述模型,因此在时间复杂度上要远远大于本文的方法,而本文中的方法简单高效,只是利用了图像的颜色信息来分割图像获得一致性区域,由一致性区域分布建立图像的描述模型。

4 结 论

提出一种新的图像分割算法:基于视觉一致性的区域分割,首先将图像量化,量化后每一个码值代表了图像中一个一致性区域,对区域中的码值定义分等级的连通性,可获取图像中区域的空间分布的拓扑结构,然后将每一个区域建立它们的描述子,然后将其用于图像检索当中。实验得出对于具有场景明确的图像,有很好的效果,对于某些图像类别的检索效果不是很好,原因之一是图像库^[10]中图像类别的划分并不是完全按照视觉一致性来划分的,而是根据具体的场景来划分,实验结果表明,对于有具体场景的图像类别的检索结果要好于场景不是很明显的图像类别的检索结果;原因之二是颜色空间维数较低,只有32维,这是需要进一步研究的方向。另外视觉一致性的区域分割为基于语义的图像检索提供了一个很好的方法,一方面它将图像分割成独立的基于视觉一致性的几个区域,这样更加有利于理解图像;另一方面,对这些独立的区域建立基于语义的描述,也变得非常简单。下一步的研究将是如何将这样的一致性分割应用到各种类型的图像当中。

参考文献 (References)

- [1] Chris Dance, Jutta Willamowski, Lixin Fan, et al. Visual categorization with bags of keypoints [C]//Workshop on Statistical Learning in Computer Vision, ECCV. Berlin, Germany: Springer, 2004:1-22.
- [2] David G Lowe. Object recognition from local scale-invariant features[C]//The Proceedings of the Seventh IEEE International Conference on Computer Vision. New York, America: IEEE, 1999:1150-1157.
- [3] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10):1615 - 1630.
- [4] Li Feifei, Perona P. A bayesian hierarchical model for learning natural scene categories[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, America: IEEE, 2005:524-531.
- [5] Li Feifei, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories[J].//Computer Vision and Image Understanding, Special issue on Generative Model Based Vision, 2007, 106(1): 59-70.
- [6] Yang Yubin, Chen Shifu, Zhang Yao. A novel image retrieval approach combining multiple features of color-connected regions [C] //Advances in Applied Artificial Intelligence, 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Berlin Heidelberg, Germany: Springer-Verlag, 2006: 510-519.
- [7] Tomasz Malisiewicz, Alexei A Efros. Improving spatial support for objects via multiple segmentations [C]//British Machine Vision Conference 2007. Warwick, England: BMVA, 2007.
- [8] Jimmy Wales. HSL and HSV [EB/OL]. (2007-06-01) [2009-02-15]. http://en.wikipedia.org/wiki/HSL_and_HSV.
- [9] Li Jia, Wang James Z. Download Databases [EB/OL]. (2003-09-01) [2009-02-15]. http://wang.ist.psu.edu/do_cs/related.
- [10] Li Jia, Wang James Z. Automatic linguistic indexing of pictures by a statistical modeling approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9):1075-1088.
- [11] Li Jia, Gray R M, Olshen R A. Multiresolution image classification by hierarchical modeling with two dimensional hidden markov models[J]. IEEE Transactions on Information Theory, 2000, 46(5):1826-41.