

文章编号: 0253-2239(2009)10-2721-05

基于文本检索技术的 CBIR 算法研究

赵 珊^{1,2} 汤永利¹

(¹ 河南理工大学 计算机科学与技术学院, 河南 焦作 454003
² 南京邮电大学江苏省图像处理与图像通信重点实验室, 江苏 南京 210003)

摘要 为了有效地利用成熟的文本检索技术,提出了一种新的图像检索算法。根据人眼的视觉特性,借鉴方块编码的思想将图像分成互不重叠的子图像块,这些子图像块在很大程度上体现了原始图像中的边缘及纹理信息,对这些子图像块进行定义,从而构造对表征图像内容有意义的图像子特征;在此基础上,把图像动态映射成文本描述形式,然后以所映射的文本为关键字,利用成熟的文本检索技术来实现图像检索。该算法不仅充分利用了图像的内容信息,而且有效地引入了文本检索技术,实验结果表明该算法具有较高的检索效率。

关键词 图像处理;基于内容的图像检索;视觉特性;方块编码;文本检索

中图分类号 TP391 **文献标识码** A **doi:** 10.3788/AOS20092910.2721

Image Retrieval Based on Text-Retrieval Technology

Zhao Shan^{1,2} Tang Yongli¹

(¹ College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454003, China
² Jiangsu Provincial Key Laboratory on Image Processing and Image Communication, Nanjing
University of Posts and Communication, Nanjing, Jiangsu 210003 China)

Abstract In order to introduce effectively text-retrieval technologies into content-based image retrieval, a novel method is presented. Firstly, according to the human visual feature and principle of the block truncation code (BTC), the image is divided into non-overlapped and equally-sized sub-blocks which can embody the texture and edge information of the original image. These blocks are defined and the sub-features are extracted, based on which the image is mapped into the text. Then, the text is chosen as key words and the technology of text-based retrieval is adopted for image retrieval. The content of the image and the text-based retrieval technology are considered, and experimental results have shown that the proposed method has sound and robust retrieval performance.

Key words image processing; content-based image retrieval (CBIR); visual feature; block truncation code; text-based retrieval

1 引 言

基于内容的图像检索 (Content-based image retrieval, CBIR) 是指利用图像自身所包含的丰富的视觉信息 (如颜色、纹理、形状及空间分布等) 来进行图像的检索^[1~4]。由于人们对图像理解和感知的局限性,在进行相似性检索时仍然存在着很多难题。

图像检索是从文本信息检索技术发展而来的,由于图像和文本两者本身特征间存在的差异性,文本检索技术中许多成功的索引技术并不能应用到图像检索中,为此,许多专家进行了深入的研究。文献[5]在进行检索时,仅仅利用文本检索模型来确定视觉特征的权值,从严格意义上来说并没有利用文本索

收稿日期: 2008-08-04; **收到修改稿日期:** 2008-11-12

基金项目: 河南省教育厅自然科学基金基础研究基金 (2008B520015, 2009B520013)、河南理工大学博士基金 (B2008-61, B2009-91) 和南京邮电大学江苏省图像处理与图像通信重点实验室开放基金 (ZK208002) 资助课题。

作者简介: 赵 珊 (1975—), 女, 博士, 副教授, 主要从事基于内容的图像检索及多媒体等方面的研究。

E-mail: zhaoshan_9228@163.com

导师简介: 周利华 (1942—), 男, 教授, 主要从事多媒体技术和信息安全等方面的研究。E-mail: zhllh@mti.xidian.edu.cn

引技术。文献[6]则采用特征矢量量化和聚类的方法定义图像的关键块,然后采用文档分析技术来实现图像检索,但由于该方法在定义图像关键块时使用的矢量量化方法不能保证各关键子块之间的独立性,并且生成码书的算法既复杂又费时,因而影响了最后的检索效果。文献[7]结合人眼的视觉特性,借鉴方块编码的思想来定义图像的关键子块,引入文本检索技术中的直方图模型来进行图像检索,虽然在一定程度上提高了图像的检索效率,但由于直方图模型在进行特征统计的时候忽略了关键块在图像中的位置信息,因而也容易造成误检和漏检。

基于此,本文提出了一种基于文本检索技术的图像检索算法。结合人眼的视觉特性,构造对表征图像内容有意义的图像子特征。并将这些子特征映射成文本描述形式,然后以所映射的文本为关键字,利用成熟的文本检索技术进行匹配检索。该算法不仅有效地提取了图像的纹理、形状等内容特征,而且将有效的文本检索技术引入到图像检索中。

2 特征提取

对每幅图像而言,不同区域都有着不同的结构特征,有的区域灰度比较均匀,没有很明显的明暗对比,而有的区域却有着很复杂的灰度差,明暗对比明显。而且,人眼对灰度变化的敏感程度跟背景有关,它随平均灰度的变化而变化,即人眼对图像细节的

分辨力与图像的灰度阶差有关。当图像本身的灰度阶差较小时,人眼的分辨力会降低,反之亦然。因此,可以将图像按照灰度阶差分成不同形状的块,来表示图像中的内容信息。

方块编码是一种有效、快速的有损数字图像压缩技术,它比矢量编码易于实现,算法简单且保留了图像的边缘信息。由于人的大脑皮层的感受细胞感受的视觉信息是有方向性的,基于这种视觉模型,我们把图像分割成 $m \times m$ 的小块,根据分块变化的形式,设计一组对人眼有意义的模块。即用一个低频分量和若干个方向的高频分量的图像块来表示图像的平坦区域和各个方向的边缘,从而来表示分块内的灰度变化,以匹配视觉的生理机制,寻求对图像准确而有效的描述。

假设 I 是一幅大小为 $M \times N$ 的灰度图像。首先将 I 划分为 $m \times m$ 大小的互不重叠的子块,对于每个子块,计算块内像素的灰度均值 μ 和平均色差 σ 。然后根据方块编码的思想,在每个子块中,灰度值大于均值 μ 的像素点赋值为 1,反之赋为 0,这样就得到了一系列 $m \times m$ 的二进制块。这些二进制块在一定程度上体现了图像中像素的分布,相似的分布会产生相同的二进制块,在算法中定义这些二进制块为图像的子特征。部分子特征的提取过程如图 1 所示(这里 $m = 2$)^[7,8]。

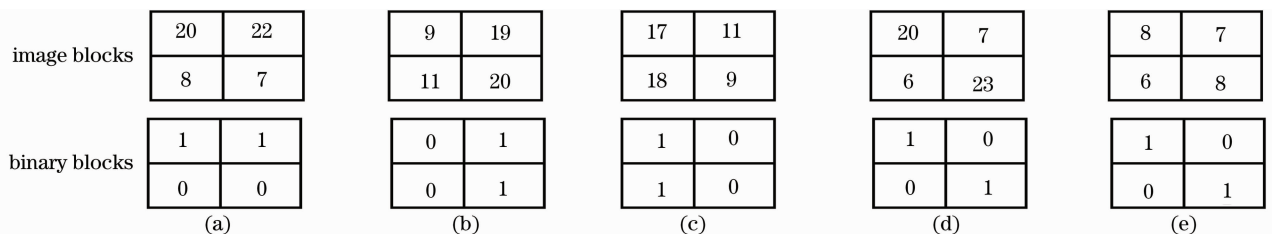


图 1 部分子特征块的提取

Fig. 1 Extraction of the sub-feature

在提取图像的子特征时,会出现如图 1(d)和图 1(e)所示的情况,针对灰度相差很大的两个图像块提取的子特征也可能会相同。因此,在算法中,设定了一个阈值 γ ,当图像块的平均色差小于这个阈值时,就把这个块看作是均匀块,四个灰度像素全设为“0”。反之,就按上述的方法提取它的子特征。采用统计的方法,经过多次实验,发现当这个阈值为整个图像平均色差的 0.0025 倍时可以取得满意的效果。在后面的实验中,取 $\gamma = 0.0025\sigma$,其中 σ 是图像的平均色差。

针对图像的子特征,我们定义一个子特征描

述符

$$I = \{ \{c_i, p_i\}, i = 1, \dots, M \}, \quad (1)$$

式中 M 为图像中总的子特征数目, c_i 为子特征相应的表示符, p_i 为子特征 c_i 在图像中所占的比例,且 $\sum p_i = 1$ 。我们设定在图像中占比例较大的子特征在描述图像内容时比较重要,因此,在构造子特征描述符时,首先将图像中占比例最大的子特征块及其比例作为第一个向量,其次是占比例第二的子特征块向量,依次类推。从定义上可以看出,针对不同种类的图像由于其表征图像的内容的不同提取的子特

征数目 M 也有可能不同。

在特征描述符的基础上,将图像映射为文本描述形式。首先我们对每个子特征块从左到右,从上到下进行扫描,以形成的二进制序列作为其索引值,然后采用一些文本中常用的字符如 A、B、C 等来对应表示相应的子特征 c_i 。算法中,用 A 表示索引值为 0 的子特征块,即均匀块, B 来表示索引值为 1 的子特征块,依次类推。最后针对每幅图像提取的特征描述符中的向量采用这些字符进行表示,形成一组字符串序列。这样,每幅图像就可以用一组字符串来进行描述,这样就完成了图像到文本的映射过程。这里的图像检索已经转化为基于文本的检索技术。计算其相似度就是计算两个字符串的匹配程度。

3 相似性度量

将图像映射为字符串后,采用文本检索技术中的字符串匹配技术进行相似性度量。根据相似学中测度相似性系统的相似度基本方法^[9,10],两个字符串的相似度 S 定义如下:

$$S(S_1, S_2) = \alpha Q_n + \beta Q_s = \alpha \cdot \frac{K}{M + N + K} + \beta \sum_{i=1}^K \lambda_i q(s_i), \quad i \in [1, k] \quad (2)$$

式中 $S_1 = \{a_1, a_2 \dots a_m\}$, $S_2 = \{b_1, b_2 \dots b_n\}$ 表示两个字符串, Q_n 是特征数目相似程度, Q_s 表示特征相似程度, α, β 分别表示 Q_n 和 Q_s 对相似元素的整体相似度的互补性,且 $\alpha, \beta \in [0, 1]$, $\alpha + \beta = 1$ 。 K 表示字符串 S_1, S_2 间相似元素的数量, M, N 分别表示字符串 S_1, S_2 中的元素数量。 λ_i 为反映相似元素对字符串相似度的影响程度的权值, $\lambda_i \in [0, 1]$, 且 $\sum_{i=1}^K \lambda_i = 1$, 定义如下:

$$\lambda_i = \left[i / \sum_{t=1}^K t + j / \sum_{l=1}^K l \right] / 2, \quad (3)$$

式中 i, j 分别表示相似元素 S_{ij} 为字符串 S_1 的第 i 个相似元, S_2 中的第 j 个相似元素。 $q(s_i)$ 表示每一相似元素相似程度大小的相似元素值,在只考虑字符串的字面特征时,也就是不考虑字符串中字符的位置信息对匹配结果的影响时, $q(s_i) = 1$, 这时就会出现 $S(ABC, BCA) = 1$ 。但是,针对本文算法提取的特征来说,位置信息是很重要的一个因素。因此,在匹配时,考虑平移代价因素,引入字符串相似元素的平移距离来对 $q(s_i)$ 进行修正。定义

$$q(s_i) = \frac{1}{1 + |i - j|}, \quad (4)$$

$|i - j|$ 表示字符串 S_1, S_2 左对齐或右对齐后,相似元素 s_{ij} 在字符串 S_1 的位置 i 与在字符串 S_2 的位置 j 的差的绝对值。

4 实验及讨论

为验证本文算法的有效性,在包含 1500 幅图像的通用图像库中进行了实验,图像库中包括了种类丰富的动物、建筑、自然景物、花卉、山脉等彩色图像。在对彩色图像处理时首先采用下式将其转化为灰度图像,然后再进行检索实验。

$$I = 0.3R + 0.59G + 0.11B, \quad (5)$$

实验运行环境为 Windows XP 操作系统,系统配置为 P41.7 GHz, 384 MB。

采用“精确度 (precision)”和“检索率 (recall)”作为相似检索的评价准则。其中检索精度定义为查询返回结果中与检索图像相关的图像占返回的图像个数的百分比,即

$$p = r/N. \quad (6)$$

检索率定义为查询结果中与检索图像相关的图像占全部相关图像的百分比,即

$$r = r/M, \quad (7)$$

式中 N 为查询返回的图像个数, r 是查询结果中与检索图像相关的图像个数, M 表示测试集合 S 中与检索图像相关的图像总数。对于不同的图像检索算法,在相同的检索条件下检索的精确度越高表明该算法的检索效果越好。

以上述评价准则为依据,进行了两组不同的对比实验。首先,针对提取子特征时不同的分块方式对算法检索性能的影响进行了实验。在进行子特征的提取时,分块方式即子块尺寸的选取直接影响到图像内容描述的有效性。大小不同的子块捕捉的尺度信息不同,较小的块体现图像的局部信息,较大的块描述相邻子块的相关信息,反映全局变化。实验中针对 $1 \times 1, 2 \times 2, 4 \times 4, 8 \times 8$ 四种分块方式的检索结果进行了实验,并采用“检索率”来对检索结果进行评价。实验中从图像库中抽取 8 类图像,每类抽取 2 幅图像作为示例图像进行 16 次检索,取这 16 次检索率的平均值作为每种分块方式最后的检索结果。采用 1×1 分块方式的平均检索率为 75.2%,采用 2×2 分块方式的平均检索率为 87.4%,采用 4×4 分块方式的平均检索率为 79.9%,采用 8×8 分块方式的平均检索率为 63.5%。可以看出,小的分块方式具有较好的检索性能,这是因为在方块编码中,小的分块引起的图像失真较小,但

是,也可以看到,1×1方式的检索性能却不如2×2,4×4,这是因为1×1分块方式丢失了图像的空间分布信息。实验结果表明2×2是最理想的一种分块方式。因此,算法中采用2×2的分块方式。

同时,将本文算法同文献[7]和文献[6]的检索性能进行了对比实验。在图像库中随机选取8类图像,每类图像抽取6幅图像共组成48次查询,取这48次检索结果精确度和检索率的平均值作为算法的平均检索结果。实验中在利用公式(2)进行相似性度量时,取 $\alpha = \beta = 0.5$,这样就赋予特征数目相似程度和特征相似程度同等重要性。图2给出了本文算法同文献[6]和文献[7]算法在“精确度”和“检索率”上的对比曲线。由实验结果可以看出本文算法的检索效率均优于其他两种算法,而且由于本文算法在提取图像的子特征时融合了人的视觉特性,检索结果更符合人类的视觉感受。同时,由于采用的字符串匹配方法中考虑了字符在字符串中的位置信息,因此和文献[7]采用直方图模型相比,检索效率要高。

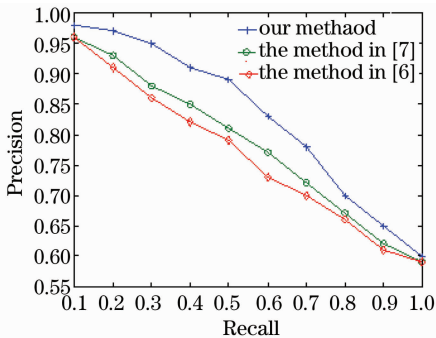


图2 三种算法检索率与准确度平均变化曲线

Fig. 2 Average recall and precision graphs for three methods

为了检验本文算法的复杂度,在相同的软硬件环境下,从特征提取时间复杂度和图像检索时间复杂度两个方面将本文算法同其它两种方法进行了比较。(1)特征提取时间复杂度:实验中在图像库中任意取100幅图像,利用本文算法提取特征所需的平均时间为35.28 s,采用文献[7]提取特征所需的平均时间为48.12 s。采用文献[6]算法提取特征所需的时间为196.53 s。(2)图像检索时间复杂度:实验采用从1000幅图像中检索出与例子图像相似的前20幅图像并显示,其中图像的索引特征已保存在不同的文件中。本文算法的平均检索时间为2.36 s,文献[7]的平均检索时间为2.59 s,文献[6]算法的平均检索时间为3.17 s。从实验结果可以看

出,本文算法的特征提取时间复杂度和检索时间复杂度都小于其他两种算法。

为了进一步检验本文算法的性能,我们采用更客观的方法ANMRR(Average normalized modified retrieval rank)^[11]进行评测。设 $T(Q_i)$ ($i = 1, 2, \dots, u$)表示图像库中与图像 Q_i 相似的所有图像数目, $t = \max\{T(Q_1), T(Q_2), \dots, T(Q_u)\}$, $l = \min\{4T(Q_i), 2t\}$, 设与例子图像相似的图像在检索结果序列中所处位置为

$$R(k) = \begin{cases} k, & k \leq l \\ l+1, & k > l \end{cases} \quad (8)$$

从而,采用ANMRR评测方法,结果A的定义如下:

$$A = \frac{1}{u} \cdot \sum_{i=1}^u \frac{\sum_{k=1}^{T(Q_i)} [R(k)/T(Q_i)] - 0.5 - 0.5T(Q_i)}{l + 0.5 - 0.5T(Q_i)} \quad (9)$$

表1给出了不同算法ANMRR的计算结果。由该表也可以看出本文方法优于其它两种方法。

表1 3种方法ANMRR计算结果

Table 1 The ANMRR result for three methods

Method	A
The proposed method	0.1365
The method in [7]	0.1852
The method in [6]	0.2378

5 结 论

图像检索是从文本检索发展而来的,但由于文本和图像二者的差异性,文本检索的许多成熟的技术难以应用到图像检索领域。本文借鉴方块编码的思想,结合人眼的视觉特性,提取图像的子特征块,在此基础上,将图像动态映射为文本描述形式,引入字符串匹配技术进行图像的相似性度量。实验结果表明,该算法具有较好的检索效果。结合其他匹配方法,诸如结构匹配方法或各种基于语义的匹配方法,从而实现对匹配结果进一步的考证与综合应用,以提高检索效率则是下一步研究的内容。

参 考 文 献

- 1 M. J. Swain, D. H. Ballard. Color indexing [J]. *Int. J. Computer Vision*, 1991, 7(1): 11~32
- 2 Zhao Shan, Cui Jiangtao, Zhou Lihua. Image retrieval based on bit-plane distribution entropy [J]. *J. Electronics & Information Technology*. 2007, 29(4): 795~79
- 3 赵 珊, 崔江涛, 周利华. 基于位平面分布熵的图像检索算法 [J]. *电子与信息学报*, 2007, 29(4): 795~79
- 3 Jiang Lijun, Luo Yongxing, Zhao Jun et al.. Haptic CT image

- retrieval based on the combination of Gabor filters and support vector machine[J]. *Chin. opt. lett.*, 2008, **6**(7): 495~498
- 4 Chaobing Huang, Quan Liu. Color image retrieval using edge and edge-spatial features [J]. *Chin. Opt. Lett.*, 2006, **4**(8): 457~459
- 5 Y. Rui, T. S. Huang, S. Mehrotra. Content-based image retrieval with relevance feedback In MARS[C]. *Proc. IEEE Int. Conf. on Image Processing*. Vol. 2, 1997. 815~818
- 6 L. Zhu, A. D. Zhang, A. B. Rao *et al.*. Keyblock: An approach for content-based image retrieval [C]. *Proc. 8th. ACM Int. Conf. on Multimedia*, 2000. 157~166
- 7 Zhao Shan, Sun Junding, Zhou Lihua. A novel image retrieval method based on keyblock[J]. *Acta Photonica Sinica*, 2007, **36**(2): 376~379
- 赵 珊,孙君顶,周利华.一种新的基于关键字块的图像检索算法[J]. *光子学报*, 2007, **36**(2): 376~379
- 8 E. J. Delp. Image compression using block truncation coding[J]. *IEEE Trans. on Comm.*, 1979, **27**: 1335~1342
- 9 Zhou ML. some concepts and mathematical consideration of similarity system theory [J]. *J. System Science and System Engineering*, 1992, **1**(1): 84~92
- 10 Zhang Chengzhi. A model for Chinese string similarity based on multi-level features[J]. *J. the China Society for Scientific and Technical Information*, 2005, **24**(6): 696~701
- 章成志.基于多层特征的字符串相似度计算模型[J]. *情报学报*, 2005, **24**(6): 696~701
- 11 H. Y. Lee, H. K. Lee. Spatial color descriptor for image retrieval and video segmentation [J]. *IEEE Trans on Multimedia*, 2003, **5**(3): 358~367