

一种基于用户感兴趣区域的图像检索算法

金磊¹, 陈优广², 严敏³

(1. 华东师范大学教育信息技术学系, 上海 200062;

2. 华东师范大学信息学院, 上海 200062;

3. 同济大学交通运输工程学院, 上海 200092)

摘要:利用图像全局特征的检索不能很好地检索用户想要的对象内容,而基于分割后各区域特征的检索又过分依赖于复杂的图像分割算法。针对上述两者的缺点,文中提出了一种基于用户感兴趣区域的图像检索算法。该算法首先对样例进行多分辨率树状分解,再由用户选择分解后的任意多个感兴趣的子图,提取子图的特征以进行相似性度量,并应用相关反馈以更好地捕获用户的检索意图。该方法无需对图像进行复杂的分割就能提取对象特征,且经由实验证明具有较高的查全率。

关键词:基于内容的图像检索;感兴趣区域;相关反馈

中图分类号:TP391.4

文献标识码:A

文章编号:1005-3751(2006)03-0104-03

An Image Retrieval Algorithm Based on Region of Interest

JIN Lei¹, CHEN You-guang², YAN Min³

(1. Department of Educational Information Technology, East China Normal University, Shanghai 200062, China;

2. Institution of Information, East China Normal University, Shanghai 200062, China;

3. Institution of Transportation Engineering, Tongji University, Shanghai 200092, China)

Abstract: Image retrieval using the feature of the entire image neglects the objects in which the user is interested, while the result of segmented regions-based retrieval excessively depends on complicated segmentation algorithm. An image retrieval algorithm based on region-of-interest is proposed in this paper aiming at covering those two shortages. This algorithm firstly decomposes the sample image into sub-images by multiresolution tree, and then the feature of the sub-images of interest selected by the user is extracted for similarity measure. We also investigate the use of feedback to better capture the user's intention. This approach can extract the feature of object without segmentation, and our experiments show that it can achieve high rate of recall.

Key words: content-based image retrieval; region-of-interest; relevance feedback

0 引言

基于内容的图像检索是利用图像本身的信息,借助现有的图像处理技术创造某种可以辨别图像特征的机制,根据每幅图像都有的可比较特征来进行检索。大量的基于内容的图像检索研究都考虑全局图像信息,然而在多数情况下,用户更关心的是图像中具有语义的区域,因此一些图像检索系统中引入了图像自动分割及自动区域提取技术。然而图像分割与人的主观认识有密切联系,到目前为止还不存在一种通用的方法,也没有一个判断分割质量的标准,因而分割结果必然会造成与人的主观认识上的差异,导致无法准确地提取相关区域的视觉特征,降低了检索结果的有效性。

文中提出一种基于用户感兴趣区域(region-of-interest, ROI)的图像检索方法,并在检索过程中应用相关反馈算法进一步学习用户的检索意图,避免了复杂的图像分割问题。实验证明,该方法简单、有效,具有较高的查全率。

terest, ROI)的图像检索方法,并在检索过程中应用相关反馈算法进一步学习用户的检索意图,避免了复杂的图像分割问题。实验证明,该方法简单、有效,具有较高的查全率。

1 用户 ROI 的定义

用户定义的 ROI 是指用户感兴趣的一个或多个经过固定划分的子图所组成的集合。最简单的固定划分子图的方法是四叉树分解^[1],其主要思想是将图像分解成 4^p 个正方形子图(p 是分解级数, $p = 1, 2, \dots$)。四叉树分解要求原图像大小必须为 $2^n \times 2^n$ 像素,文中采用一种伪四叉树分解法,它只要求原图像大小为 $2^{n-1}m \times 2^n$ 或 $2^n \times 2^{n-1}m$, $m \geq 2$,降低了对图像大小的约束条件,在执行上后者与前者的差别主要在第一级分解上,即第 1 级分解时,横向子图数为 $\frac{\text{图像的宽}}{\text{图像宽和高的最大公约数}}$,纵向子图数为 $\frac{\text{图像的高}}{\text{图像宽和高的最大公约数}}$,从第 2 级开始,只需对上一

收稿日期:2005-06-26

作者简介:金磊(1981—),女,上海人,硕士研究生,研究方向为视觉检索。

级的每个子图分别应用标准四叉树分解。当 $m = 2$ 时,伪四叉树分解就退化为四叉树分解。

检索开始时用户首先确定分解级数,系统根据级数对样例做固定划分,然后用户在划分后的样例上选择所有感兴趣的子图,最终 ROI 就是这些子图的集合。通常用户在对图像库检索时不仅需要检索在视觉上相似的区域,还希望多个区域在空间位置上满足一定的约束条件,而这一约束条件已经隐含在 ROI 中。例如:想要检索“海上日出”,显然,用户会按照这样的原则选择 ROI:样例图像中包含蓝色区域(大海)、黄色区域(太阳)的子图,且黄色区域在蓝色区域上方。通过定义 ROI,实际可以获得的特征除了 ROI 的颜色和纹理,还包括 ROI 内各子图间的相对位置信息。对 ROI 定义相应的范围框(Bounding Box),它是包含了所有 ROI 的最小矩形框,如图 1 所示。

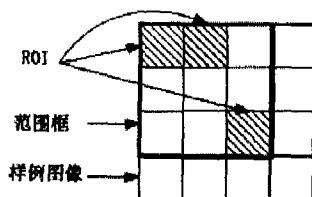


图1 ROI与范围框

下一节将详细说明如何利用范围框提取和匹配图像特征。

2 基于 ROI 的图像检索算法

2.1 特征描述

考虑到对原始图像分解后的各子图的特征描述,文中改进了 MARS 检索系统中的图像特征模型^[2],设图像特征表示模型为 $F(I, H, R)$,各参数分别定义如下:

① I = 原始图像

② $H = \{h_i\}$ 是原始图像经分解后的区域集合, $i = 1, 2, \dots, N_r, N_r$ 为分解后的子图数。

③ $R = \{r_{ij}\}$ 是图像各区域的低层次视觉特征的集合, r_{ij} 是特征向量,表示图像区域 h_i 的特征, $r_{ij} = [r_{ij1}, r_{ij2}, \dots, r_{ijk}, \dots, r_{ijk}]$ 。

此特征表示模型是一个通用模型,可以适用于任何的颜色、纹理、形状等特征向量。

2.2 特征相似性度量

为了更精确地表达用户对不同特征和区域的关注度,需定义各特征和区域的权重,通过计算加权特征向量间的欧氏距离来度量两幅图像的相似性。对应图像特征表示模型 F ,定义权重模型 $W = [W_{ROI}, W_R, W_r]$, W 的各成分说明如下:

① $W_{ROI} = [W_{ROI1}, W_{ROI2}, \dots, W_{ROI1}, \dots, W_{ROI L}]$, $W_{ROI} \in [0, 1]$, L 为 ROI 中区域的数目。 W_{ROI} 表示 ROI 中各区域的权值。

② $W_R = [W_1, W_2, \dots, W_j, \dots, W_J]$, $W_j \in [0, 1]$, J 等于 F 中的特征数。 W_R 表示图像各特征,如颜色、纹理、

形状等的权值。

③ $W_r = [W_{r1}, W_{r2}, \dots, W_{rj}, \dots, W_{rJ}]$, 其中 $W_{rj} = [W_{rj1}, W_{rj2}, \dots, W_{rjk}, \dots, W_{rjK}]$, $W_{rjk} \in [0, 1]$ 。 W_r 表示特征向量内各分量的权值。

检索时,对图像库中每一幅图像都定义一个与范围框同样大小的活动“窗口”,称之为范围窗口。令活动窗口在图像库中的每一幅图像上做从左向右、从上到下的平移,对于每一个平移位置,称当前范围窗口内与 ROI 内子图相对位置一致的那些子图为窗口映射。这样,基于 ROI 的图像检索就是要寻找含有与 ROI 最相似的窗口映射的 N 幅图像。图 2 显示了范围窗口平移时形成的不同的窗口映射, (X, Y) 表示 ROI 的左上第一个子图在分解网格中处于横向第 X 、纵向第 Y 个位置。 (X_i, Y_i) 表示第 i 个平移位置处,窗口映射中左上第一个子图在分解网格中所处的位置。

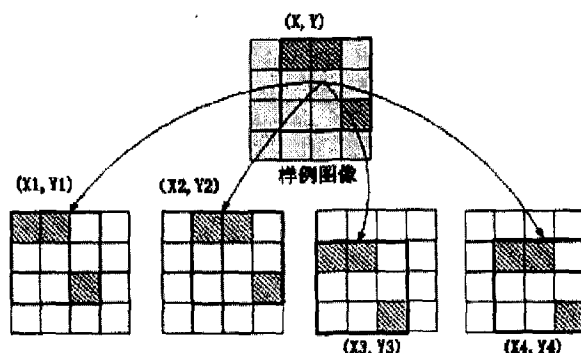


图2 范围窗口在一图库图像上的窗口映射

设 Q_p 是用户在图像 p 级分解时选取的 ROI, L 为 ROI 的子图数,则 $Q_p = \bigcup_{p=1}^L Q_{p_p} = \{Q_{p1}, \dots, Q_{pL}\}$, 且令 B 表示窗口映射, N_t 表示范围窗口在一幅图像内部可以平移的次数, $N_t = [\text{横向子图数} - (\text{Max}(X_i) - \text{Min}(X_i) + 1) + 1] \times [\text{纵向子图数} - (\text{Max}(Y_i) - \text{Min}(Y_i) + 1) + 1]$, 则样例 I 与数据库中任一图像 I' 的相似性度量定义为:

$$d(I, I') = \min_{1 \leq t \leq N_t} (d(Q_p, B_t(I'))) \quad (1)$$

其中, $B_t(I')$ 是范围窗口在 I' 中第 t 次平移后的窗口映射, $d(Q_p, B_t(I'))$ 是 Q_p 与 $B_t(I')$ 中各对应子图的特征向量的加权欧氏距离之和。式(1)的意义是:样例图像 I 与数据库中任一图像 I' 的相似性,是由 ROI 分别与 I' 的 N_t 个窗口映射的距离的最小值来表示的,取得最小距离的窗口映射即为 I' 与 ROI 的最接近匹配。

根据图像特征模型, $H = \bigcup_{i=1}^N h_i$ 和 $H' = \bigcup_{i=1}^N h'_i$ 分别为 I 与 I' 的子图集,建立下标集合 $\text{Sub} = \{i_l | h_{i_l} \in Q_p, 1 \leq l \leq L\}$, $\text{Sub}' = \{i'_l | h'_{i'_l} \in B_t(I'), 1 \leq l \leq L\}$, 以向量欧氏距离作为度量函数, $d(Q_p, B_t(I'))$ 的计算公式为:

$$d(Q_p, B_t(I')) = \sum_{i_l \in \text{Sub}, i'_l \in \text{Sub}'} W_{ROI} \left[\sum_{j=1}^J W_j \left[\sum_{k=1}^K W_{rjk} (r_{ijk} - r'_{ijk})^2 \right]^{1/2} \right] \quad (2)$$

初次检索时,各权值初始化为: $W_{ROI} = 1/L$, $W_j =$

$1/J, W_{rjk} = 1/K$ 。遍历图像库中所有图像, 利用式(2), (1)度量与 ROI 的相似性, 取度量值最小的前 N 幅图像作为检索结果。

3 相关反馈

相关反馈是通过用户对检索返回的结果进行评价和标记, 并将标记的“相关”和“不相关”的信息作为训练样本反馈给系统进行学习, 指导下一轮检索, 从而使检索结果更好地接近语义层次的要求^[3]。文中改进了 Rui 的基于权重调整的相关反馈算法^[2], 加入了区域权重的调整算法和针对“不相关”图像的惩罚算法。令“相关”和“不相关”图像分别称为“正例”和“负例”, 根据用户标记, 对每一幅结果图像赋 Score 值, 正例图像的 Score 值为 1, 负例图像 Score 值为 0, 权重调整算法如下:

算法 1 调整 $W_R = [W_1, W_2, \dots, W_j, \dots, W_J]$

设根据总体特征相似性得到的 N 个最相似图像是 $RT = \{RT_1, RT_2, \dots, RT_N\}$, 根据单个特征相似性得到的 N 个最相似图像是 $RT^j = \{RT_1^j, RT_2^j, \dots, RT_N^j\}$, 则 RT 中的相关图像与 RT^j 重合越多, W_j 就越大, 具体算法是:

$$W_j = \begin{cases} W_j + \text{Score}_i, & RT_i^j \in RT \\ W_j + 0, & RT_i^j \notin RT \end{cases}, i = 1, 2, \dots, N$$

然后对 W_j 进行归一化处理, $W_j = \frac{W_j}{\sum_{j=1}^J W_j}$ 。

算法 2 调整 $W_r = [W_{r1}, W_{r2}, \dots, W_{rj}, \dots, W_{rJ}]$

令正例集合 $U = (u_1, u_2, \dots, u_{n_{\text{pos}}}, \dots, u_{N_{\text{pos}}})$, 负例集合 $V = (v_1, v_2, \dots, v_{n_{\text{neg}}}, \dots, v_{N_{\text{neg}}})$, $N_{\text{pos}}, N_{\text{neg}}$ 分别为 U, V 中的图像数, 那么 W_{rjk} 的改变依赖于 RT 中正例图像的 ROI 的特征向量。设 ROI 数为 $L, H = N_{\text{pos}} \times L$, 对于 U 中图像的特征 r_{ij} , 组成特征矩阵 $[r_{ij}^n]_{H \times K}$, 矩阵元素 r_{ij}^n 表示正例图像 $u_{n_{\text{pos}}}$ 的第 h_i 个子图的特征 j 的第 k 个分量, 其中 $i_l \in \text{Sub}^n$ 。

计算矩阵 $[r_{ij}^n]_{H \times K}$ 的每列元素的标准差 σ_{jk} , 该值越小, 说明特征的这个分量重要性越大, W_{rjk} 取为标准差 σ_{jk} 的倒数,

$$W_{rjk} = \frac{1}{\sigma_{jk}}, \text{再对 } W_{rjk} \text{ 归一化, } W_{rjk} = \frac{W_{rjk}}{\sum_{k=1}^K W_{rjk}}。$$

算法 3 调整 $W_{\text{ROI}} = [W_{\text{ROI}1}, W_{\text{ROI}2}, \dots, W_{\text{ROI}L}, \dots, W_{\text{ROI}L}]$

对样例各 ROI 的每一区域与 RT 中的 $u_{n_{\text{pos}}}$ 一一计算区域间特征向量距离 d_{ij}^n :

$$d_{ij}^n = W_{\text{ROI}} \left\{ \sum_{j=1}^J W_j \left[\sum_{k=1}^K W_{rjk} (r_{ijk} - r_{ij}^n)^2 \right]^{1/2} \right\}$$

组成矩阵 $[d_{ij}^n]_{N_{\text{pos}} \times L}$, 计算每一列标准差 $\sigma_{\text{ROI}l}$, 新的

$$W_{\text{ROI}l} \text{ 取为 } W_{\text{ROI}l} = \frac{1}{\sigma_{\text{ROI}l}}。$$

将第 $w-1$ 次检索结果中的正反馈图像集自动置于第 w 次的检索结果中, 也就是 $RT^w = [u_1^{w-1}, u_2^{w-1}, \dots, u_{N_{\text{pos}}^{w-1}}^{w-1}, RT_1^w, RT_2^w, \dots, RT_{N-N_{\text{pos}}^{w-1}}^w]$, 这样系统始终能保持 $N_{\text{pos}}^w \geq N_{\text{pos}}^{w-1}$ (N_{pos}^w 是第 w 次反馈的正例图像)。

令 d_v 是数据库图像 I_i 与第 w 轮检索时反馈的 V 中各负例图像的距离, 定义惩罚函数

$$\text{Pun}^w(d_i) = \begin{cases} 0 & w = 1 \\ \text{Pun}^{w-1}(d_i) + \sum_{v \in V} 1/d_v^w & w = 2, 3, \dots \end{cases} \quad (3)$$

上式中, $i = 1, 2, \dots, M - \sum_{j=1}^{w-1} N_{\text{neg}}^j$, M 为数据库中图像的总数, N_{neg}^j 为每一轮返回的负例图像数。每一轮反馈后的惩罚因子不断累加, 则与任一轮反馈的负例图像非常接近的那些图像, 将不再可能作为结果返回给用户, 即惩罚函数使检索结果“不像”负例图像^[4]。考虑惩罚因子的作用, 第 w 轮检索后的 I_i 与样例图像的距离度量表达式为:

$$\hat{d}_i^w = d_i^w + \text{Pun}^w(d_i), i = 1, 2, \dots, M - \sum_{j=1}^{w-1} N_{\text{neg}}^j \quad (4)$$

4 实验结果

笔者利用 MATLAB6.5 实现了一个图像检索实验系统。采用的图像特征是颜色直方图^[5]和 Gabor 纹理特征^[6]。实验图库由 40 类, 每类 100 张的 COREL 标准图片组成。以一张建筑图像作为样例, ROI 为经 2 级分解后的含“双塔”的子图, 图 3 显示了检索结果。

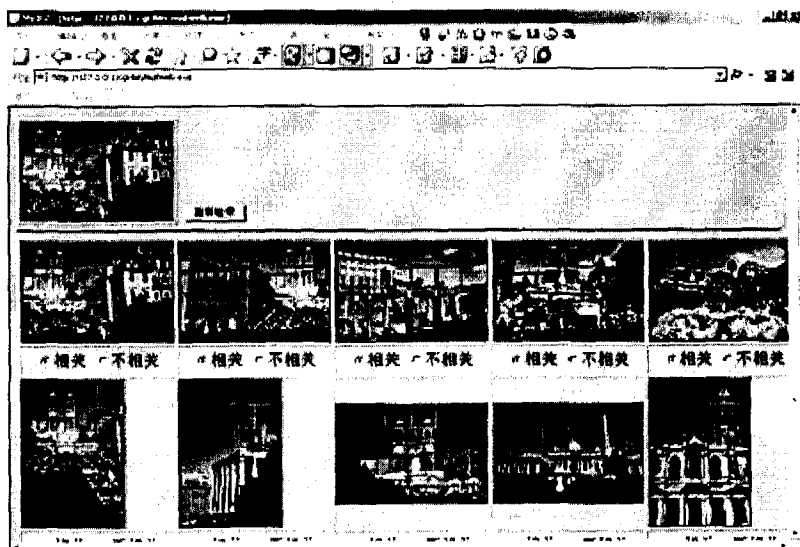


图 3 含“双塔”的建筑图像检索示例

将所有建筑类图像按其结构特征分成 4 类, 统计不同反馈次数下的平均查全率, 从图 4 的统计数据中可以看出, 在系统经过 8~10 次的反馈之后查全率能够从 50%~60% 提高到 80%~90%, 可见相关反馈对提高检索准确度有比较明显的作用。

(下转第 109 页)

乘模型去拟合 X 和 Y 所引进的误差。

在理想的情况下, X 中的误差的来源与 Y 中的误差来源完全相同, 即影响 X 与 Y 的因素相同。但实际上, X 中的误差与 Y 中的误差并不相同。因而, $t \neq u$, 但当两个矩阵同时用于确定因子时, 则 X 和 Y 的因子具有式(9)的关系:

$$u = bu + e \quad (9)$$

式中 b 所表征的即 u 和 t 间的内在关系。

2.3 分类结果和方法稳定性的评价方法

评价分类结果和方法的稳定性, 具体有以下步骤:

(1) 随机从 N 个样本数据集中抽取 N_1 个样本数据作为训练集, 这 N_1 个样本将用来建模, 剩余的 $N_2 = N - N_1$ 个样本将作为测试集用来验证模型。

(2) 使用训练集建模, 并使用 leave-out-one 对模型进行交叉验证。

(3) 使用剩余样本对模型进行验证, 这可以有效地防止过拟合现象。

(4) 为了进一步防止过拟合现象, 重复 1~3 步来验证模型。

3 局限和问题

由于 SVM 找到的是全局最优解, 因此在很多问题上比其它统计学习技术更有优越性。但是作为一种尚未成熟的新技术, SVM 也存在着以下局限:

(1) 如何提高 SVM 的训练性能。由于 SVM 训练法在本质上是解决一个二次规划(QP)问题, 如果训练集规模很大, SVM 的训练性能将大大降低。可从两个方面来思考此问题: 其一, 从训练算法角度入手; 其二, 从缩小训练集入手。由于支持向量仅仅是类与类边缘的一小部分样本, 离支持向量较远的样本对选择支持向量的影响较小, 于是可以用其它方法(比如聚类、去噪等)来大大缩小训练

样本, 从而提高 SVM 训练性能。

(2) 如何选择核函数及核函数的参数。由于核函数确定后, 用户只能对参数进行调整, 因此核的选择对 SVM 的性能至关重要。如何针对具体问题选择最佳的核仍是个难以解决的问题。

(3) SVM 主要是解决两分类问题, 但对用 SVM 来进行多分类及回归问题的性能有待进一步研究^[8]。

参考文献:

- [1] Cristianini N, Shawe-Taylor J. 支持向量机导论[M]. 李国正, 王猛, 曾华军译. 北京: 电子工业出版社, 2004.
- [2] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 16-18.
- [3] Poggio T. An overview of statistical learning theory[J]. IEEE Trans Neural Networks, 1999, 10(5): 35-45.
- [4] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods[M]. Cambridge, England: Cambridge University Press, 2000.
- [5] Joachims T. Making large-scale SVM Learning Practical Advances in Kernel Methods - Support Vector Learning[M]. America: MIT Press, 1999.
- [6] Molla M, Waddell M, Page D, et al. Using Machine Learning to Design and Interpret Gene-Expression Microarray[J]. AI Magazine, 2004, 25: 32-35.
- [7] Dubitzky W, Granzow M, Berrar D. Data Mining and Machine Learning Methods for Microarray Analysis[A]. In: Lin S M, Johnson K F. Methods of Microarray Data Analysis - Papers from CAMDA 2000[C]. Boston, Kluwer: Academic Publishers, 2001.
- [8] Cho S, Won H. Machine Learning in DNA Microarray Analysis for Cancer Classification[M]. APBC, Cambridge, England: Cambridge University Press, 2003. 189-198.

(上接第 106 页)

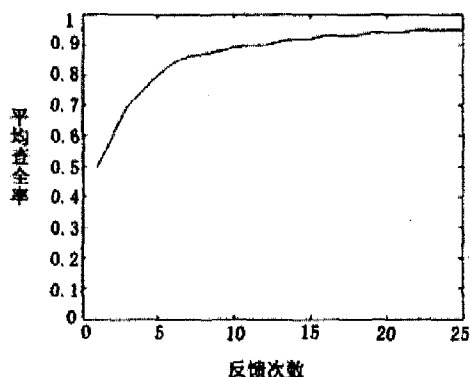


图4 建筑类图像在不同反馈次数下的平均查全率

参考文献:

- [1] Malki J, Boujemaa N, Nastar C, et al. Region Queries without Segmentation for Image Retrieval by Content[A]. Proceedings of the Third International Conference on Visual Information

and Information Systems[C]. London, UK: Springer-Verlag, 1999. 115-122.

- [2] Rui Y, Huang T S, Mehrotra S, et al. Relevance feedback: A Power tool for interactive content-based image retrieval[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998, 8(5): 644-655.
- [3] 陈纯. 计算机图像处理技术与算法[M]. 北京: 清华大学出版社, 2003.
- [4] Su Zhong, Zhang Hongjiang, Li S, et al. Relevance Feedback in Content-Based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressive Learning[J]. IEEE Transactions on Image Processing, 2003, 12(8): 924-936.
- [5] 邢伟利. 图像检索中颜色的特征提取及匹配算法[J]. 微机发展, 2004, 14(2): 86-88.
- [6] Manjunath B S, Ma W Y. Texture Features for Browsing and Retrieval of Image Data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 837-842.