

基于分块权重和相关反馈技术改进 Zernike 矩描述符的执行效率

冯玉才 吴潇 梁俊杰

(华中科技大学计算机科学与技术学院, 武汉 430074)

摘要 Zernike 矩是一种基于区域的形状描述符,它适合于描述具有复杂边界的目标。原始的 Zernike 矩描述符认为形状中任何位置的像素都具有相同的重要性,基于此,提出了一种改进的 Zernike 矩描述符。它首先采用预定义的两个半径值对原始形状进行分块,提取各分块的 Zernike 矩值作为图像的形状特征向量,然后采用相关反馈信息和各分块间的距离方差来动态调整各个分块的权值系数,根据欧式距离来计算图像间的相似度。改进的 Zernike 矩描述符可以根据人类视觉特征灵活确定形状各部分的重要性,而且相关反馈技术使得提取出来的图像更加接近用户的需求。实验结果表明,该方法能够有效地改善基于形状特征的检索效果。

关键词 图像检索 形状特征 Zernike 矩 权重 相关反馈

中图分类号: TP3 文献标识码: A 文章编号: 1006-8961(2007)06-1086-05

Improving Retrieval Performance of Zernike Moment Descriptor via Weighted Partitions and Relevance Feedback

FENG Yu-cai, WU Xiao, LIANG Jun-jie

(Computer Science and Technology Department, Huazhong University of Science and Technology, Wuhan 430074)

Abstract Zernike moments are used as a shape descriptor for complex shapes that are difficult to be defined with a single contour such as trademarks. The Zernike moments of a given shape are calculated as correlation values of the shape with Zernike basis functions in that all the pixels of the shape regardless of their positions contribute with the same weight to the Zernike moments. The proposed modified Zernike Moment descriptor is obtained by the following two steps: firstly divide the original shape into three parts of inner, middle and outer regions with two predetermined radius, then calculate the Zernike moment of each part separately. The modified descriptor takes account of the partition radius of the shape according to human perception, meanwhile, using relevance feedback technology to fix the importance of the each part as mentioned above could improve the efficiency of retrieval process. Euclidean distance is used to compute the distance between two shapes. Experimentation under various test conditions shows the effectiveness of the proposed modified method.

Keywords image retrieval, shape representation, Zernike moment, weight, relevance feedback

1 引言

基于内容的图像检索 (content-based image retrieval, CBIR) 是指利用图像本身的视觉内容 (例如, 颜色、形状、纹理、对象的空间关系等) 来检索与例子图像 (query image) 相似的数据库图像。如何实现快速有效的图像检索, 关键在于采用何种特征作

为索引以及特征如何匹配, 这是基于内容的图像检索技术的核心问题。

形状是图像最显著的核心特征之一, 也是人类视觉系统进行物体识别时所识别的关键信息之一。基于形状的图像检索, 是基于内容的图像检索的一个十分重要的方面。形状描述符主要有基于边界和基于区域两大类, 前者只利用形状的外部边缘, 而后者利用形状的全部区域。

基于区域的特征提取同时利用了形状边界和内部像素,它对于噪声和形状失真具有较好的健壮性,因此更加适用于描述普通的形状。自 20 世纪 60 年代提出起,基于区域的力矩描述符的使用就非常盛行。这里包括几何矩^[1]、不变矩^[1]、Legendre 矩^[2]、Zernike 矩^[2]和 Chebyshev 矩。矩可以描述目标区域的灰度分布特性和边界形状,由于充分利用了形状内部和边界的大量信息,所以能更全面地反映目标本质特征,它还对旋转具有不变性。

正交矩具有最小的信息冗余,对图像的表达和描述能力好,因此在形状描述领域得到了广泛的应用。Teague^[2]提出了两种不同的连续正交矩,Zernike 矩和 Legendre 矩,分别基于正交的 Zernike 和 Legendre 多项式。许多研究表明,Zernike 比 Legendre 更有优越性,因为它具有较好的特征表示能力和较低的噪声敏感度。

原始的 Zernike 矩描述符认为形状中任何位置的像素都具有相同的重要性,但在实际应用中,人们对形状中不同区域的关注程度一般不同。因此根据人类的视觉感知特性,提出了一种改进的 Zernike 矩描述符,实验结果表明,它能有效地改善基于形状特征的检索效果。

2 Zernike 矩

Zernike 矩不变量只具有旋转不变性,为了使它同时具有平移和缩放不变性,首先使用 Leu^[3]提出的图像规格化方法对形状进行预处理,并使之映射到一个紧凑的单位圆内,圆心即为形状的质心。函数 $f(x, y)$ 的 Zernike 矩 Z_{nm} 的定义如下:

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} V_{nm}^*(x, y) f(x, y) dx dy \quad (1)$$

$$V_{nm}(x, y) = R_{nm}(x, y) \exp(jm \tan^{-1} \frac{y}{x}) \quad (2)$$

其中, $V_{nm}(x, y)$ 表示 n 阶 Zernike 矩的基函数, $x, y \in [0, 1]$, n 为非负整数, m 为满足 $n - |m|$ 为偶数且 $|m| \leq n$ 的整数。

实值多项式表示如下:

$$R_{nm}(x, y) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} (x^2 + y^2)^{\frac{n-s}{2}} \quad (3)$$

在数字图像域,Zernike 矩的表达形式如下:

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y V_{nm}^*(x, y) f(x, y) \quad (4)$$

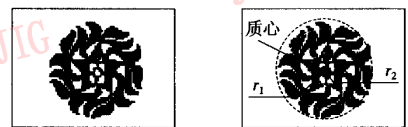
$$x^2 + y^2 \leq 1$$

$|Z_{nm}|$ 被称为 Zernike 矩不变量。它有以下优点:旋转不变性,Zernike 矩的模对于旋转不变;冗余度小,由于变换基是正交的,因此信息表达的冗余性小,所提取特征的相关性和冗余性小^[2];可表示性,容易构造高阶不变矩,可提供更多特征用于识别;健壮性,鲁棒性好,抗噪能力强^[4]。

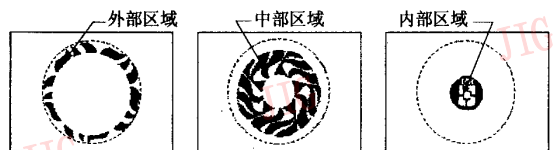
3 改进的 Zernike 矩描述符

原始的 Zernike 矩不变量认为,形状中任何位置的像素对 Zernike 矩都具有相同的权重。根据实际经验,用户对图像不同区域的关注程度是不相同的,有的注重目标的外部形状,有的注重目标的内部区域。本文提出的改进算法基本思想是,首先将预处理后的形状按照已定义的两个半径进行划分,然后提取每个分块中的 Zernike 矩作为其形状特征向量,检索阶段先逐一计算出对应分块之间的距离,然后进行某种程度的加权累加作为两幅图像之间的实际距离。两幅图像之间的距离越小,则它们越相似。给定两幅图像 p 和 q , 设 q 为查询对象, p 为图像数据库中的任意一幅图像。

(1) 根据预定义的两个半径值 r_1 和 r_2 ($0 < r_1 < r_2 < 1$), 将原始形状划分为内部、中部和外部 3 个区域。设其分块集合为 $T = \{T_{inner}, T_{middle}, T_{outer}\}$, 如图 1 所示。



(a) 原始形状 (b) 预处理后的形状



(c) 划分后的外部 (d) 划分后的中部 (e) 划分后的内部

图 1 形状划分示例

Fig. 1 Shape partition

(2) 分别提取 3 块区域的 Zernike 矩, 构成图像的特征向量。

(3) 计算两个形状的相似度。首先计算对应分块之间的距离,然后进行加权累加,得到图像 p 与 q 之间的距离。计算公式如下:

$$D_{ij}(p, q) = \|F_{ij}(p) - F_{ij}(q)\|, i \in T \quad (5)$$

$$D_i(p, q) = \sum_{j \in N_i} D_{ij}(p, q) \quad (6)$$

$$Dist(p, q) = \sum_{i \in T} (\omega_i \cdot D_i(p, q)) \quad (7)$$

$$\omega_i \in [0, 1], \sum_{i \in T} \omega_i = 1 \quad (8)$$

其中, $F_{ij}(\cdot)$ 表示形状中分块 i 的第 j 个 Zernike 矩; $D_{ij}(p, q)$ 表示图像 p 和 q 之间对应分块 i 的第 j 个 Zernike 矩向量之间的距离; N_i 表示分块 i 包含的 Zernike 矩数量; $D_i(p, q)$ 表示图像 p 和 q 之间对应分块 i 之间的距离; ω_i 则为分块 i 的权重系数。对用户关注的区域,修改权重系数,提高该分块 Zernike 矩的重要性,使其在相似性度量中起更大的作用。

(4) 确定权重系数。各个区域的权重系数,开始时由用户根据实际情况设定,不同的权重系数组合,可以实现不同类型的检索:当 ω_{inner} (ω_{middle} 或者 ω_{outer}) 取值为 1,另外两个权重系数均为 0 时,图像的相似性度量只需要考虑形状内部区域(中部或者外部区域)即可,这种模糊匹配的方法可以提高检索的速度;当 3 部分权重系数相等时,检索效果与原始 Zernike 矩方法相同。

4 相关反馈技术

相关反馈是一种非常有效的反馈方法。用户根据先前检索结果,通过调整权重来给检索系统提供更多更直接的信息,从而使系统能更好地满足用户的需求^[5]。利用相关反馈技术来动态调整权重系数,具体步骤如下:首先定义权重系数的缺省初始值, $\omega_{inner} = \omega_{middle} = \omega_{outer} = 1/3$;然后提取出最相似的前 N 幅结果图像,用户根据自身的查询要求和视觉感受给出反馈信息;最后量化反馈信息,通过调整分块的相似度来调整权重系数。

设 k 表示相应图像与目标样例的相似度,由用户指定每幅图的 k 值, $k \in \{0, 1, \dots, 9\}$, 值越大,相应的图像与目标样例越相似。新的查询均值向量 \bar{F}_{ij} 和对应分块间的距离方差 σ_i 为

$$\bar{F}_{ij} = \frac{\sum_{t=1}^N (k_t \times F_{ij}(t))}{\sum_{t=1}^N k_t}, i \in T \quad (9)$$

$$\sigma_i = \frac{1}{3} \sum_{j \in N_i} \sum_{t=1}^N (F_{ij}(t) - \bar{F}_{ij})^2, i \in T \quad (10)$$

其中, k_t 表示反馈过程中用户给出的图像 t 与样例图像的相似程度; $F_{ij}(t)$ 表示图像 t 中分块 i 的第 j 个 Zernike 矩特征向量。

距离方差越大表明分块间的相似程度越小,因此必须在搜索过程中降低该分块的重要性。距离测量公式中每个分块之间的权重系数如下:

$$\omega_i = \frac{1/\sigma_i}{\sum_{j \in T} 1/\sigma_j} \quad (11)$$

5 实验与评价

基于上述思想,实现了一个原型系统——DMIR,并设计了两个实验。第 1 个实验是与原始 Zernike 描述符以及不变矩 IM 描述符的检索性能对比实验,以便比较分块方法的查询效果;第 2 个实验则是测试本文所采用的相关反馈策略的有效性。

实验 1

第 1 个实验所采用的测试图像数据库是由 418 幅二值商标图像组成,根据人类的视觉相似性,将其划分为 12 类,分属不同的主题内容。每个类别所包含的商标图像数目从 8 到 118 不等,其中有 4 个类别的图像数目超过了 40 幅。图 2 显示了这 12 个类别中的一些示例图像。图 3 为 DMIR 系统的某次检索结果,其中工作区左下角的图像为查询样例,右边查询结果的相似度自左至右、自上向下递减。

对检索效果的评价主要使用查准率 (Precision) 和查全率 (Recall) 两个指标。查准率的含义是在一次查询过程中,系统返回的查询结果中相关图像的数目占所有返回图像数目的比例,是检索结果精确性的度量准则;查全率则指系统返回的查询结果中相关图像的数目占图像库中所有相关图像数目(包括返回的和没有返回的)的比例,是对检索健壮性的度量准则。

为了使实验更加客观,首先不考虑相关反馈的信息,每次查询中各分块的划分半径以及权重系数均由用户设定,且规定:判断两幅图像是否相似的依据是它们是否属于同一类别。对于不变矩 IM,提取形状的 7 个 Hu 矩不变量作为特征向量;对于原始 Zernike 矩方法中则提取形状的前 36 个 Zernike 矩作为特征向量;对于改进的 Zernike 矩方法中,分别提取形状各分块区域中前 20 个 Zernike 矩作为特征



图 2 不同主题类别中的示例商标图像
Fig. 2 Sample logo of different subjects



图 3 DMIR 系统的检索结果样图
Fig. 3 Retrieval result of DMIR system

向量。实验过程中,分别从 12 个商标类别中各随机抽取 50% 的图像作为查询图像,这样构成了 209 次查询。对每次查询,分别计算出查询结果的查准率和查全率,这样可以得到每个类别的平均查准率和平均查全率,进而根据各个类别所含图像数目进行加权平均,最终得到算法总的查准率和查全率。

图 4 清晰地显示了随着返回图像数目的增加,漏检目标也随之增多,各描述符的查准率均呈下降

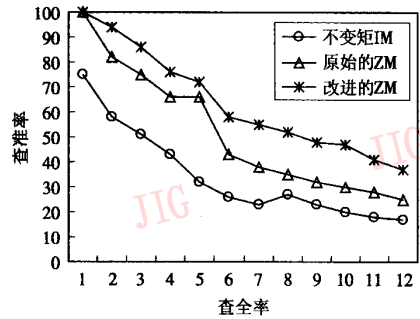


图 4 3 种形状描述符的查准率
Fig. 4 Precision of shape descriptors

趋势,但改进的 Zernike 矩描述符的检索性能始终最高。

表 1 显示了不变矩 IM、原始 Zernike 矩描述符和改进 Zernike 矩描述符的查全率。与不变矩 IM 相比,原始的 Zernike 矩描述符提高了将近 25% 的检索精确度;而改进的 Zernike 矩描述符相对于原始 Zernike 矩而言,又将检索精确度提高了将近 11%。在上述 12 类商标中,改进的 Zernike 矩对其中 7 类商标的检索结果明显优于原始 Zernike 矩。某些特殊情况例如类别 7,3 种矩的检索精度都非常的低,这可能是由于当初根据视觉相似性对图片进行分类而产生的误差引起的。

表 1 3 种形状描述符的查全率
Tab. 1 Recall of shape descriptors

类别	所含图像的数目	不变矩 IM	原始的 Zernike 矩	改进的 Zernike 矩
1	24	22.251	32.939	31.113
2	118	20.556	63.658	78.966
3	15	36.568	61.570	43.787
4	20	39.222	69.445	63.889
5	8	23.750	43.752	56.256
6	12	25.449	57.143	69.306
7	18	24.445	26.667	32.231
8	86	30.612	42.857	67.762
9	35	56.368	64.267	64.267
10	42	17.253	28.531	48.257
11	15	13.920	30.720	35.842
12	25	42.314	67.769	65.636
总数	418	25.758	51.214	62.543

实验 2

测试图像数据库的图片与实验 1 所使用的相同,图像题材包括三角形、方形、椭圆形、星形、多边形、圆形、长条形、横标以及各种花形。图 5 为 DMIR 的初次检索结果,在查询结果的前 20 幅图像中,有 3 幅图像与查询目标比较类似。图 6 是经过 1 次相关反馈后的检索结果,其中相关的图像数目增加到了 5 幅,而且相关图像的位置也更加靠前。在第 1 次相关反馈的基础上再次进行相关反馈操作得到了如图 7 所示的检索结果,可以看出,相关图像的数目已经增加到 11 幅,并且前 10 幅图像全部是相关的。这说明:通过相关反馈能比较有效地捕捉用户的查询意图,反馈的结果会朝着有利于用户要求的方向发展。



图 5 DMIR 的初次检索结果
Fig. 5 Results of original retrieval



图 6 经过 1 次相关反馈后的结果
Fig. 6 Results after first feedback



图 7 经过两次相关反馈后的结果
Fig. 7 Results after second feedback

6 结 论

提出了一种基于权重的分块 Zernike 矩描述符,它采用相关反馈信息和各分块间的距离方差来调整权重系数,有效地改善了基于形状特征的检索效果。实验结果显示了改进的 Zernike 矩描述符的有效性,且引入相关反馈信息后的查询结果更加准确,系统的检索性能明显提高。

今后,将进一步对 DMIR 检索系统的功能进行扩展,主要工作为:(1)本文所讨论的相关反馈主要是联机反馈分析,今后将进一步融入脱机反馈分析,以便在查询过程中更好地实现系统的自适应和自学习;(2)研究有效的索引技术,以便系统在面临大规模数据时还能有较好的查询效率。

参考文献 (References)

- 1 Hu M K. Visual pattern recognition by moment invariants[J]. IRE Transactions on Information Theory, 1962, 8(3): 179 ~ 187.
- 2 Teague M R. Image analysis via the general theory of moments[J]. Journal of Optical Society of America, 1980, 70(8): 920 ~ 930.
- 3 Leu J G. Shape normalization through compacting [J]. Pattern Recognition Letters, 1989, 10(6): 243 ~ 250.
- 4 Kim H, Kim J. Region-based shape descriptor invariant to rotation, scale and translation[J]. Signal Processing: Image Communication, 2000, 16(5): 87 ~ 93.
- 5 Salton G, McGill M J. Introduction to modern information retrieval [M]. New York: McGrawHill, 1983.