

基于对象形状的形象查询技术*

邬浩 潘云鹤 庄越挺 杨宇艇

(浙江大学人工智能研究所 杭州 310027)

E-mail: wuhao@huawei.com.cn

摘要 基于图象内容的查询是根据图象实体(或区域)的颜色、形状、纹理、空间关系等特征属性来查询图象。它把图象处理、图象识别、图象数据库3个领域的技术成果结合起来,是一个有前途的发展方向。本文所作的工作是研究和实现根据图象实体(或区域)的形状来查询图象。内容包括:(1)人对于形状的认知过程;(2)表征形状的特征量集合;(3)一种快速有效的图象匹配算法;(4)原型系统 Photo Engine。

关键词 形状,基于图象内容查询,图象数据库,内角。

中图法分类号 TP392

如何存储和管理大量的计算机图象是图象数据库技术一直在努力解决的问题。由于图象数据和文本数据有着本质上的区别,在文本数据领域得以成功应用的传统数据库技术如果一成不变地搬到图象数据领域来,其结果必然是低效甚至无效的,因而需要研究针对图象的表示存储等技术。从另一方面看,传统数据库的许多成果,如 SQL 语言、索引技术等也是值得大力借鉴的。上述两方面的结合是图象数据库技术现今发展的主流。

近年来,国外在图象数据库技术的研究中出现了一个新热点——基于图象内容的查询 QBIC(query by image content)。^[1,2]所谓基于图象内容查询,是指根据图象中物体(或区域)的颜色、形状、纹理等特征以及这些特征的组合来查询图象。象 IBM 公司的 Almadier 研究中心、密西根大学计算机系、新加坡国立大学等国外高校和科研机构都纷纷对这个新热点展开了研究,并取得了一定成果。这一新的研究热点涉及计算机图象处理(Image Processing)、计算机图象识别(Image Recognition)和图象数据库技术(Image Database),把这3个领域的研究成果进行有效结合是该技术取得成功的基础。这项技术充分体现了图象的信息特点,又充分结合了传统数据库技术,是一项在理论研究和实际应用中都很有前途的新技术。特别是,随着近年来 Internet 应用热潮的掀起,随着各个行业对 IDB 技术的新的要求的不断涌现,基于图象内容的查询技术的应用前景变得更为诱人。

目前,基于图象内容的查询主要有3个研究方向:(1)基于对象(区域)颜色的查询;(2)基于对象(区域)形状的查询;(3)基于对象(区域)纹理的查询。本文选定其中的一个研究方向——基于对象(区域)形状的查询作为研究目标,主要包括两方面的研究内容:

(1) 找到足够的表征物体形状的特征参数

由于应用的对象是拥有大量图象的图象数据库,因此,要求这些特征量应该足够简单,并且能够事先计算和存储,以便于查询。在选取特征参数时还应该注意到图象中的物体有如下特点:(a)其尺寸大小不是统一的;(b)其方向不是一致的;(c)在图象中的相对位置是不一致的,如图1所示。因此,特征参数的值应该不受或很少受物体的尺寸、方向和相对位置这3个因素的影响,这样才能较好地匹配。事实上,只要某个特征参数对其中任何一个因素能保持恒定不变,就应该是可选特征参数。

(2) 快速而有效的匹配算法

首先,查询的结果是一系列的图象,按照相似的程度依次排列;其次,由于它的应用对象是大的图象数据库,必须保证一定的速度。因此,该匹配算法与图象识别中的匹配算法有所不同,它不要求精确的匹配。在必要的时候,可以在

* 本文研究得到国家自然科学基金、国家 863 高科技项目基金、国家攻关计划基金和国家教委博士点基金资助。作者邬浩,1971 年生,工程师,主要研究领域为图象处理与模式识别,多媒体信息检索。潘云鹤,1946 年生,教授,博士生导师,主要研究领域为人工智能, CAD/CG, 认知科学。庄越挺,1966 年生,副教授,主要研究领域为人工智能, ICAD, 图形图象处理, 多媒体信息检索。杨宇艇,1972 年生,硕士,主要研究领域为图形图象数据库,多媒体信息检索。

本文通讯联系人:邬浩,深圳 518057,深圳华为公司交换业务部终端研究部

本文 1997-03-31 收到原稿,1997-05-29 收到修改稿

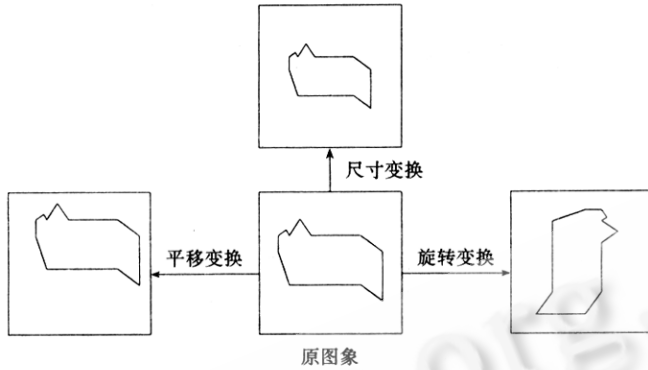


图1 同一形状在不同图象中的不同表现

精度和速度之间作一个折衷。

1 人对形状的认知过程

形状是人类视觉中一个十分重要的特征,在选择表征形状的特征量之前,有必要先从认知过程的角度来考察一下人类视觉对形状的认识。

尽管人类图象识别的原理至今还不清楚,但有一个可接受的假设:人的视觉的形状识别是分层特征识别。在比较两个物体形状的时候,首先获得的是对这两个物体的比较模糊的描述,比如:物体甲很狭长而物体乙很扁平;物体甲很圆而物体乙很尖锐等等,这些特征都属于高层的特征。利用这些特征通常可以区分两个形状差别很大的物体。当高层特征完不成识别时,就需要选取低层特征加以比较。这些特征包括:(1) 物体的欧拉特性,包括物体孔的个数、物体由几个分立的部分组成等等;(2) 物体的分解特性,比如一把桌子可以分解为桌面、桌腿;(3) 物体尖角的多少;(4) 物体凹点的多少和凹陷的程度等。

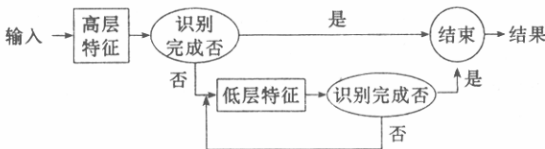


图2 识别过程示意图

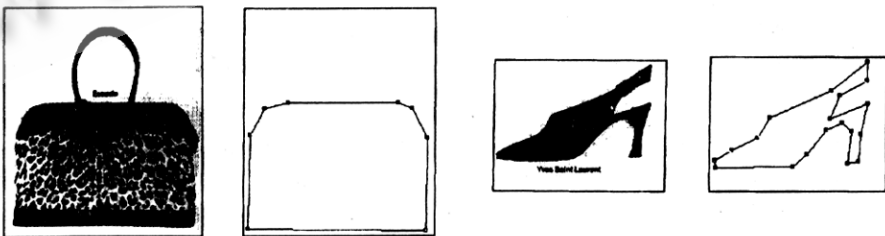
一个反馈修正的过程,如图2所示。

2 形状的表征

表征形状的方法主要有以下几种:

- 本征方程 本征方程 $\psi(s)$ 是斜率(或曲率)对弧长(或弦长)的函数,它是平移不变的,归一化后是比例尺不变的,也可使它对旋转不变。^[3]Freeman 给出了离散化的本征方程,这就是链码表示。^[4]对于闭轮廓,其特点是它的本征方程为一周期函数。

- 几何常量 如圆形度、离心率等。这些几何常量通常对平移、比例、旋转都不变,但它们在有些情况下不同形状的物体也会得出相同的值。图3是不同圆形度的两个例子。



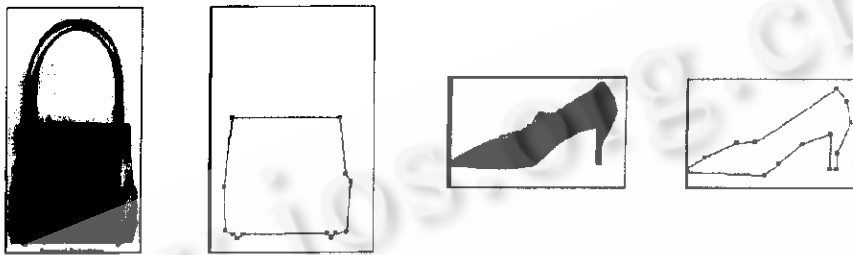
圆形度 = 7.3

圆形度 = 55.8

图3

• Fourier 级数 Fourier 系数的低频分量对噪声不敏感,为表示轮廓提供了一种非完备,但却有用的总体描述。Fourier 描述子的主要缺点是轮廓上感兴趣的部分,如有无遮挡,由于映射到全部系数中而看不到了。这通常可用轮廓分段的方法加以改进。

• 力矩 设 μ_{pq} 是图象的任意 $(p+q)$ 次中心力矩,把 μ_{pq} 适当组合,能得到许多对物体平移、尺寸放缩或旋转不变的矩常量,根据这些矩常量能有效地识别物体。其中应用得最广泛的矩常量是“绝对矩常量集”,它共有 7 个矩常量,都是 μ_{pq} 的非线性组合。矩常量与 Fourier 描绘子有同样的缺陷,即对图象的描述太过笼统,难以反映细节。图 4 是两个不同的二阶矩常量的例子。



二阶矩常量=1.0

二阶矩常量=13.7

图 4

除了上述方法以外,还有一些其它方法,如利用曲率、拐点、凹点等等特征来识别物体。

在上述方法的基础上,我们提出了如下一些新的形状特征量。

(1) 狭长度

通过坐标原点,斜率为 θ 的直线 $y=x\tan\theta$ 周围的图象 f 的惯性力矩 M_θ 用下式定义。

$$M_\theta = \sum_{i=1}^N \sum_{j=1}^M (j\cos\theta - i\sin\theta)^2 f(i, j) = \mu_{02}\cos^2\theta + \mu_{20}\sin^2\theta - \mu_{11}\sin 2\theta$$

由 $dM_\theta/d\theta=0$ 可求得,使 M_θ 为最大值 $M_{\theta_{max}}$ 和最小值 $M_{\theta_{min}}$ 。我们定义

$$\text{狭长度} = \frac{M_{\theta_{max}}}{M_{\theta_{min}}}$$

狭长度这个特征量能反应物体的狭长程度,其值越大表示物体越狭长。如图 3 所示的两个形状,正方形的狭长度是 1,这是狭长度的下限值,它说明正方形对其重心是完全对称的;而长方形的狭长度大于 1,且长方形越扁,狭长度的值越大(如图 5 所示)。

(2) 基于内角的特征量

内角在多边形形状识别中有着极为重要的作用^[5],下面我们将从多边形的内角入手,推出一些表征多边形形状的特征量。

多边形的顶点数 多边形的顶点越多,表示形状越复杂。对于两个顶点数相差悬殊的多边形,有理由认为其形状相差很大。图 6 是不同顶点数的多边形的例子。

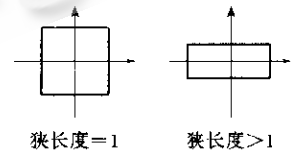
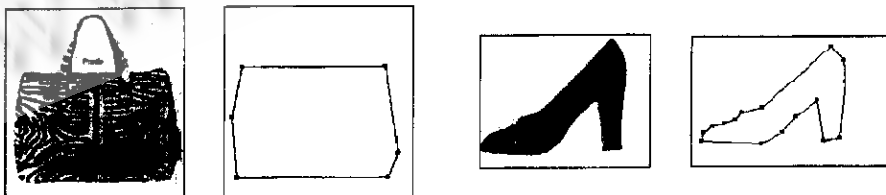


图 5



顶点数=6

顶点数=15

图 6

多边形内角的平均值 内角平均值是内角特征的一个总体描述,它能区分诸如三角形和正方形等形状。图 7 是两个例子。

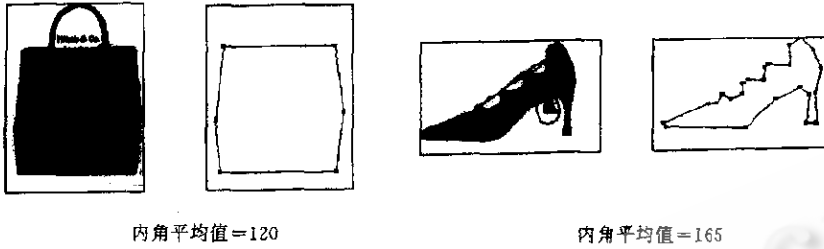


图7

多边形内角的标准方差 内角标准方差是对内角分布情况的一个总体描述,多边形越规则,其值越小,因此,它能区分诸如正方形和不规则四边形等形状。

多边形向凹点 凹点反映在内角上就是内角大于180°的点,凹点的有无分别与多边形的两大类——凹多边形和凸多边形相对应,是非常重要的形状特征,利用内角可以统计多边形的凹点数目,这个特征量可以用来区分有着明显差别的形状,如图8所示。

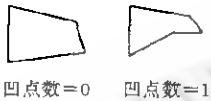


图8

多边形内角的直方图 所谓多边形内角的直方图,就是将0~360°划分为等距或不等距的k个区间,然后分别统计落在这k个区间上的内角数目,内角直方图描述的是多边形内角的分布特性,如图9(a)所示的内角直方图在大于180°的角度区间上没有分布,表明该多边形是一个凸多边形;而图9(b)所示的内角直方图在大于180°的角度区间上有分布,表明该多边形是一个凹多边形。

内角区间分隔点的选取决定着内角直方图的表述结果,因此,分隔点选取的好坏直接影响着内角直方图描述内角分布特性的能力,由于一个多边形的顶点通常是几十个,因此,我们选取4~5个角度区间是比较合适的,也就是说,要选取3~4个区间分隔点,作为凹凸性的临界点,180°是一个很好的区间分隔点;另外,我们还选取了90°和270°作为另外两个分隔点,前者是锐角和钝角的临界点,后者是前者在180°~360°区间上的对应。

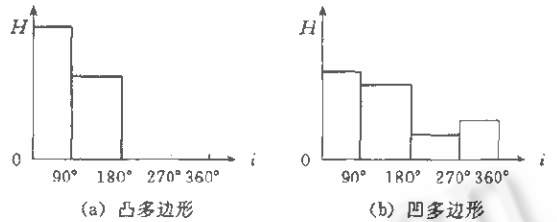


图9

图10给出了内角直方图的两个具体例子,为方便图象匹配时使用,我们对内角直方图作标准化处理,即用内角概率密度函数 $P_i(i)$ 来表示。

根据前面的研究结论,我们选定如下一些形状特征量:

- 圆形度 = $4\pi \frac{\text{(面积)}}{\text{(周长)} * \text{(狭长度)}}$
- 狭长度 = $\frac{M_{\theta_{max}}}{M_{\theta_{min}}}$
- 惯性主轴方向 $\theta_0 = \frac{1}{2} \tan^{-1} \frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \pm \frac{n\pi}{2}$
- 二阶矩常量 $M_1 = \mu_{20} + \mu_{02}$

μ_{pq} 是大小为 $N \times M$ 的图象 $f(i, j)$ 的 $(p+q)$ 次中心力矩,其定义如下:

$$\mu_{pq} = \sum_{i=1}^N \sum_{j=0}^M (i - i_g)^p (j - j_g)^q f(i, j), \quad (i_g, j_g) \text{ 为物体重心}$$

- 多边形的顶点数
- 内角平均值 $a_0 = \sum_{i=1}^N a_i / n$
- 内角标准方差 $\delta = \sqrt{\sum_{i=1}^N (a_i - a_0)^2}$
- 多边形的凹点数
- 内角概率函数 $P_i(i)$, 其中 $i = 1, 2, 3, 4$, 它们对应的内角区间分别是 $[0 \sim 90^\circ)$, $[90 \sim 180^\circ)$, $[180 \sim 270^\circ)$,

[270~360°].

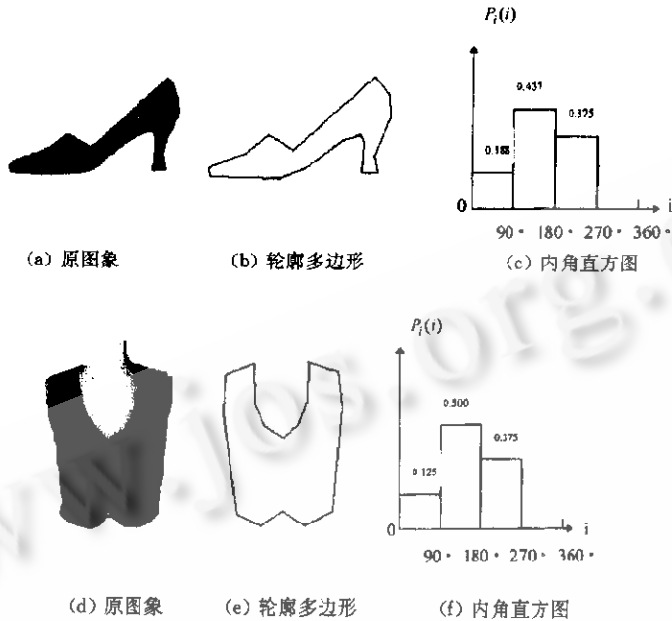


图 10

上述这些特征量中,绝大多数都与物体的方向、位置、尺度大小无关,少数的如惯性主轴方向虽然与物体的方向相关,但因为它是物体的重要特征,所以能把图象很好地分类,并且在某些查询中被用户所强调。

3 匹配算法

由于本文所面对的实际上是一个通用的形状识别问题,在选取匹配算法的时候必须考虑如下一些特点。

(1) 不再象普通的模式识别问题一样存在固定的几类模式,只需把识别对象判定为这几个类中的一个。在本文中,每一个图象对象就是一个模式,对应着特征向量空间中的某个点,查询模板同样也对应着特征向量空间中的某个点,现在需要判定的是,哪些图象对象在特征向量空间中所对应的点与查询模板所对应的点距离最小(相似度最大)。

(2) 一般来说,图象数据库中的图象是很多的,如果需要计算每个图象实体与查询模板之间的距离,其计算量之大通常是难以接受的。因此,必须先对图象实体进行筛选,把那些很不相似的图象实体先行除去,以保证查询速度。

基于这样的考虑,在进行全面的形状特征量比较之前,必须选用一些有效的形状特征量,首先把形状相差很大的图象从匹配队列中清除出去,以保证系统运行的效率。因此,采用以下查询策略,首先把所有图象对象都放入匹配队列中,进而选取查询模板的某些形状特征值,与匹配队列中图象对象的这些特征值进行比较。如果它们相差很大,则可认为该图象对象与查询模板的形状是相差很大的,从而可以很快地把它从匹配队列中删去。例如,描述物体形状的多边形顶点数是一个非常重要的形状特征。一般来说,其值越大形状越复杂,越小则形状越简单。因此,如果图象数据库中的某个图象对象的顶点数与查询模板的顶点数相差很大(比如相差两倍以上),那么我们认为它与查询模板在形状上很不相似,把它从查询结果中排除。除了多边形顶点数之外,还有其他一些特征量也可用于该操作,如狭长度和圆弧度。选用多少个特征量作这样的操作,则取决于匹配队列中图象数目的多少,少于一定的数目(例如小于 50)则停止这个操作,否则,选择下一个特征量继续进行。

把匹配队列中的图象对象限定到一定的数目后,我们运用相似性测度函数来计算查询模板与匹配队列中每一个图象对象的形状相似程度,用 0~1 之间的一个数来表示。这个总的相似性又由各个特征量的相似性经过加权求和而得,各个特征量的相似性函数定义如下:

对于那些是单个数字的形状特征量,查询模板和图象对象在该特征量上的相似度是

$$S_i = 1 - \frac{|\text{查询模板的特征值} - \text{图象对象的特征值}|}{\max\{\text{查询模板的特征值}, \text{图象对象的特征值}\}}$$

其中 i 表示第 i 个特征量, S_i 的值都在 $0 \sim 1$ 的范围之内.

对于内角直方图或内角概率密度 $P_i(i)$, 相似度是

$$S_H = \sum_i \min\{\text{查询模板的 } P_i(i), \text{图象对象的 } P_i(i)\}, \text{ 或}$$

$$S_H = 1 - \sum_i |\text{查询模板的 } P_i(i) - \text{图象对象的 } P_i(i)|$$

而总的相似度是

$$S = \frac{\sum_i W_i S_i + W_H S_H}{\sum_i W_i}$$

其中 W_i 和 W_H 分别是第 i 个特征量和内角直方图的权值. 这些权值可以由我们根据形状特征量的重要程度赋予固定的值, 也可由用户根据查询要求中他所强调的重点所在加以更改.

匹配队列中的每个图象对象与查询模板的相似度计算出来后, 我们把它们按降序排列, 即最相似的图象对象排在最前面. 然后选定一个阈值, 相似值高于这个阈值的图象对象就作为查询结果输出. 阈值可以是固定的, 也可以根据查询结果的情况调整.

这样, 我们最后得到的查询结果是一批与查询模板最相似的图象对象, 输出时按照相似程度由大至小排列. 整个查询过程如图 11 所示.

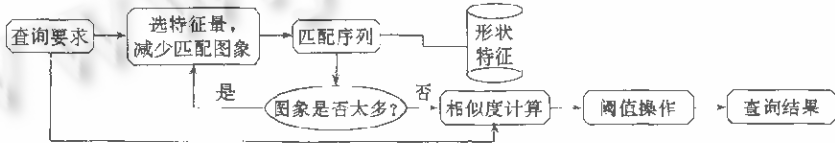


图11 基于形状特征的查询

4 原型系统 Photo Engine

在前面理论研究的基础上, 我们开发了原型系统 Photo Engine, 成功地应用了理论研究的成果. 系统分为两个主要模块: 图象入库模块和图象查询模块.

图象入库模块主要包括 3 个子过程: 预处理、边界提取和特征量计算, 其功能分别是:

(1) 预处理

预处理子过程主要完成如下功能: 除去图象中的噪声, 把图象的尺寸规整化, 在需要的时候使用边缘增强算法突出图象中物体的边界, 为下一步边界提取作准备.

(2) 边界提取

目前还不存在一种通用而有效的图象对象边界提取算法, 大多数算法都难以得到完整且清晰的边界轮廓, 而且对于那些背景复杂的彩色图象, 这些算法就更容易失败. 基于这样的考虑, 我们选用手工勾边的方法提取边界.

边界提取子过程使得用户能够交互地用多边形来逼近物体的边界, 包括多边形顶点的增加、删除、调整等功能, 采用多个标志量实现这几个状态之间的转移.

(3) 特征量计算

图象中物体的边界多边形提取出来以后, 首先计算多边形的面积、周长、各个一阶和二阶力矩、重心位置、各个内角, 进而求得我们所选取的形状特征量集中的各特征量的值. 经过上述 3 个操作过程, 另外再输入一些文本描述, 包括源文件名、物体的名字、物体的特点等.

图象查询模块提供两种基于形状特征的查询途径, 根据例子的查询和根据草图的查询. 前者是以现有的某个图象对象作为查询模板, 而后者则把用户所画的形状草图作为查询模板.

该系统所建立的图象数据库中包括 300 余个图象对象, 有服装、鞋类、箱包等等.

5 小结与展望

总的看来, 本文的研究工作可小结如下:

(1) 对形状的代表方法作了比较系统的研究. 首先考察了人在比较两个形状时的认知过程, 这是一个探讨性的研

究,接着用几何表达和抽象表达模型系统地介绍了各种形状表征方法和它们之间的联系,在概述了几种主要的表征方法后,详细介绍了基于力矩的形状特征量和基于多边形内角的形状特征量,在前者中提出了“狭长度”这个特征量,在后者中提出了“内角直方图”这个表征方法,这些是有创新性的研究成果。

(2) 研究了用于图象数据库的匹配技术,提出了一种快速的基于特征量的匹配技术并加以实现。

(3) 实现原型系统 Photo Engine,该系统是在前面所述的理论研究的基础上完成的;成功地应用了理论研究的成果建立了一个模型图象数据库,实现了图象的入库和查询功能,提供了根据例子和根据草图这两种基于形状特征的查询方法。

我们的工作还可以在如下几个方面加以改进和完善:

(1) 更为方便的图象对象的边界提取方法。简单方便的提取边界的方法能大大减小图象入库时的工作量,可以考虑的方法是采用自动的或半自动的提取方法。

(2) 研究有效的形状表征方法。

如前所述,我们的研究是图象处理、图象识别与图象数据库3个研究领域的交叉点,而图象识别则是最困难的。困难表现在现有的研究一般是针对特定的应用领域,即要求识别的模式只有很少的几类,其特征量的选取因而也就有着明显的针对性,而我们则对通用性的要求非常高,需要比较任意两个形状的相似程度,这实际上是一个通用的形状识别问题,无疑是很困难的。

总之,根据图象内容的查询技术充分体现了图象信息的特点,体现了人类对计算机图象信息的根本要求,是一个非常前途的研究方向。

参考文献

- 1 Niblack *et al.* The QBIC project: querying image by content using color, texture, and shape. *International Society for Optical Engineering*, 1993, 1908: 173~187
- 2 Flicker *et al.* Query by image and video content: the QBIC system. *IEEE Computer*, 1995, 9(28): 23~31
- 3 Low. *Introductory computer vision and image processing chap.* London: McGRAW-HILL Book Company, 1991
- 4 王积分,张新荣. 计算机图象识别. 北京:中国铁道出版社,1988
(Wang Ji-fen, Zhang Xin-rong. *Computer image recognition.* Beijing: China Railway Publishing House, 1988)
- 5 帕夫利迪斯. 结构模式识别. 上海:上海科学技术文献出版社,1981
(Pafledis. *Structural pattern recognition.* Shanghai: Shanghai Science and Technology Literature Publishing House, 1981)

Query Image by Object Shape

WU Hao PAN Yun-he ZHUANG Yue-ting YANG Yu-ting

(Artificial Intelligence Institute Zhejiang University Hangzhou 310027)

Abstract QBIC (query by image content) is a method to query image by characteristics of the object (or region) of the image, which include color, shape texture and spatial relations. This new research project combines technique of IP (image process), IR (image recognition) and IDB, is a very prospecting domain. The authors deal with the investigation and realization of querying images by the shape of image objects (or region) in this paper, which includes: (1) the cognitive process of shape; (2) a set of characteristic to describe shape of image object; (3) an efficient match algorithm for image querying; (4) a prototype system for querying image by the shape of image objects, which named Photo Engine.

Key words Shape, query by image content, image database, intra-angle.