

基于压缩域的图象检索技术研究进展

黄翔宇 章毓晋

(清华大学电子工程系, 北京 100084)

摘要 压缩标准的出现,使得图象数据格式普遍为压缩格式,从而促进了压缩域内图象检索技术的迅速发展.为了使人们对基于压缩域的图象检索技术有一概括了解,该文对目前的压缩域检索技术进行了回顾和讨论:首先,介绍了图象检索系统的基本概念;然后,分类分析了不同压缩域的检索技术,包括变换域方法(如傅立叶变换、离散余弦变换(DCT)以及子带和小波变换)和空域方法(如矢量量化和分形等);接着,对这些检索方法进行了讨论和比较,并得出一些有用的结论;还举例介绍了基于压缩域图象检索技术的实际应用;最后对压缩域图象检索技术的研究发展及其应用前景指出了一些可能的方向.

关键词 计算机图象处理(520·6040) 图象检索技术 压缩域 离散余弦变换 小波 矢量量化 分形
中图法分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2003)05-0499-10

A Review of Image Retrieval Techniques in the Compressed Domain

HUANG Xiang-yu, ZHANG Yu-jin

(Department of Electronic Engineering, Tsinghua University, Beijing 100084)

Abstract The advent of compression standards, such as JPEG and MPEG, has led to the popularity of the compressed form of image data, and that has brought on the proliferation of image retrieval techniques in the compressed domain. In this paper, we make a comprehensive review and discussions on the compressed domain retrieval techniques proposed in the literature, including the latest achievements in this field. First, we give the general concepts of the image retrieval technology. Secondly, we analyze different retrieval techniques, including transform domain techniques using Fourier transform, discrete cosine transform, subbands and wavelets, and spatial domain techniques using vector quantization and fractals. Thirdly, we discuss and compare these image retrieval techniques and draw some useful conclusions. In addition, an example application of image retrieval in the compressed domain is presented. Finally, we make a discussion on some open problems and point out possible directions for further research.

Keywords Computer image processing, Image retrieval techniques, Compressed domain, Discrete cosine transform(DCT), Wavelet, Vector quantization, Fractal

0 前言

随着超大规模集成技术(VLSI、宽带网络技术如ATM、ISDN)以及图象/视频压缩标准(JPEG、MPEG)的发展,数字图象检索技术日益显示出重要性.在不久的将来,图象/视频检索可应用于多媒体信息系统、数字图书馆、电影行业、视频点播等多方面,而现在采用的图象检索方式仍普遍为基于图象

内容的检索方式.由于基于内容的图象检索能够让人们根据自己的描述,或能根据用户提出的样本图象,从庞大的图象数据库中检索到需要的图象,以达到有效合理地组织和管理整个信息系统的目的,而且这种查询方式符合人们对客观事物的感知和认识,因而便于信息系统管理的计算机化.

压缩标准的出现,使得图象数据格式普遍为压缩格式.由于传统的图象数据处理方式是在原始数据域或者是在解码以后的压缩数据上进行处理,而

压缩域的图象数据处理则是直接在压缩数据上进行处理^[1],因此基于压缩域的图象检索与传统的基于原始域的图象检索相比,有如下优点:

(1)部分省略了解压缩和再压缩的附加环节,因而对减少系统的总体处理时间非常有利;

(2)压缩域上的数据量比原始域上的数据量大大减少,因而对应用系统的总体处理效率的提高是个有利因素,尤其当系统要求实时性时,在压缩域上进行处理常是唯一的解决办法;

(3)某些压缩算法在一定程度上适应了图象检索的要求,因而对提高检索效率有帮助。

1 基于内容的图象检索系统模型

在以往的研究中,人们提出了一些基于内容的图象检索系统.这些图象归档/检索系统结构示意图见图 1^[2].这些检索系统在进行检索时,首先计算出每幅图象的多维特征向量,然后根据特征向量的相似度来进行检索.由于对多种特征的解释和量化是模糊不清的,所以检索的重点更多地放在相似度的计算上.在索引应用中,特征的选择常考虑下列性能参数:(1)区别不同图象的能力;(2)一次查询可得到的最多图象数;(3)计算和比较特征所需要的计算量^[3].据了解,目前许多视觉检索技术都是基于颜色、纹理、形状以及直方图等来直接从图象像素中抽取特征进行检索.

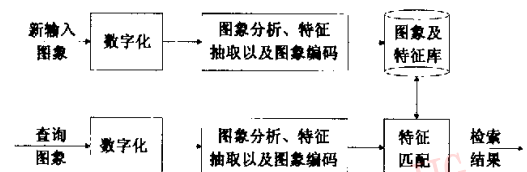


图 1 图象归档-检索系统示意图

另一方面,由于近来人们又提出了图象和视频压缩标准如 JPEG、MPEG、H. 26L 等,以用于减少带宽和存储空间,因此现在的图象和视频普遍是以压缩形式进行存储的,这就促进了许多压缩域图象检索技术的发展.压缩域图象检索区别于传统的基于原始数据域的图象检索,由于它是直接面向“数据量大大减少了”的压缩数据而进行的检索,因此它不涉及解压缩、再压缩的环节(见图 2)^[3].

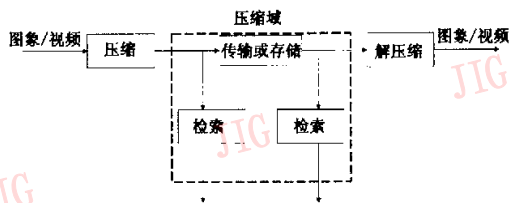


图 2 压缩域检索系统框图

因为视觉数据的海量性使得压缩技术的使用势在必行,所以,未来的多媒体数据库都将会以压缩形式存储视觉数据.为了避免对数据解压之后再行像素域的图象检索,故采取直接在压缩域上进行检索的方法,将会更为有效.在本文的讨论中,所考虑的压缩域图象检索主要是根据压缩数据来进行的,因为这些技术计算和存储索引需要的计算量小、存储空间少.根据图象编码中采取的压缩方法的不同,压缩域的图象检索技术大致上可以分为基于变换域的和基于空域的两类.其中,基于变换域的图象检索主要基于离散傅立叶变换(DFT)、KL 变换(Karhunen-Loeve Transform, KLT)、离散余弦变换(DCT)以及小波变换;基于空域的图象检索则包括 VQ(矢量量化)和分形等(见图 3).

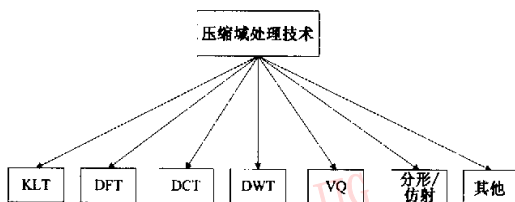


图 3 多种压缩域上基于内容的图象检索

2 基于压缩域图象检索的发展现状

2.1 离散傅立叶变换

傅立叶变换在图象处理和信号处理中有着非常重要的地位.由于 DFT 利用复数指数函数进行变换,其能量压缩特性良好,从而可提供良好的编码性能.实践证明,DFT 的某些性质在检索和模式匹配中是很有用的:首先,DFT 系数的幅度具有平移不变性;其次,空域上的相关度也可以通过 DFT 系数来计算^[4].

Stone 和 Li 提出了一种傅立叶域的图象检索

算法^[4]。这种算法提供两个阈值,用户可通过这两个阈值来独立地调整匹配的近似度。其阈值之一用于控制亮度匹配,另一个阈值用于控制纹理匹配。另外,由于这两个阈值之间是相关的,故利用傅立叶系数可以有效地计算出来,且在傅立叶系数大部分是零的时候,这种计算尤其有效率。

Augustejin 利用纹理测量对卫星图象进行分类^[5]。这种统计性的纹理测量是基于图象的傅立叶频谱幅度,其中包括:(1)最大幅度值;(2)平均幅度;(3)幅度的能量;(4)傅立叶系数幅度的方差,而且,作者还研究了基于傅立叶系数半径(即幅度)和相位分布特性的检索性能。据研究,一般傅立叶系数半径的分布对纹理的粗细很敏感,而相位的分布则对纹理的方向很敏感。据观察,当图象的频谱具有一些典型的频率时,不仅傅立叶系数半径和相位测量有着很好的分类性能,而且这种统计性的纹理测量在缺乏典型频率时,分类性能良好。

Celantano 和 Lecce 分析了傅立叶系数相位分布在图象检索中的性能^[6]。他们先对图象进行低通滤波的预处理,然后进行傅立叶变换;接着,用一个旋转的矢量在 180°的范围内,对 DFT 频谱进行扫描;最后,对每个傅立叶系数相位,计算图象成分的个数,即得到相位直方图。在计算相位直方图时,由于中频部分代表了视觉上最重要的图象成分,所以作者只考虑中频部分的相位。据研究,像素域的平移不会影响相位直方图的分布,而像素域的旋转则会引起相位直方图的循环移位。

2.2 Karhunen-Loeve 变换(KL 变换)

KL 变换是基于图象统计特性上的一种变换。据研究,KL 变换提供了最大的能量压缩,而且在统计上是最优的。由于 KL 变换的变换函数是与图象本身相关的,所以,通过比较 KL 变换域上图象的 KL 变换系数,可以获得良好的图象检索性能,但是,因为 KL 变换的计算复杂度庞大,所以实际应用中,很少有图象用 KL 变换进行压缩,也就是说,虽然在理论上,基于 KL 变换域的图象检索很有意义,但由于缺乏用 KL 变换压缩的图象,故基于 KL 变换域的图象检索并不是一种实用的方法。

Pentland 将 KL 变换用于人脸识别^[7]。文献^[7]在进行人脸识别时,首先从图象库中随机抽取部分图象来生成一系列最优的基本图象,例如本征图象;然后将查询图象映射到本征图象上;最后通过比较查询图象和目标图象的 KL 变换系数之间的欧氏距

离来识别人脸。既然 KL 变换基本图象按本征值排序,那么用较少的 KL 变换系数(15~20 个)就能表示显著的图象特征。

虽然将图象投影到 K-L 空间可以抽取它最显著的特征,但是,本征值可能也代表了成像过程的某些特征(如光照方向),而这些特征不仅与识别无关,而且增加特征值的个数并不一定能提高识别成功率。为了解决这个问题,Swets 和 Weng 提出了一种鉴别 KL 映射(DKL)的方法^[8],这样在做完 KL 变换之后,再进行鉴别分析,就可生成一组最具鉴别力的特征(MDFs)。由于 DKL 映射能保证类间距离最大化,同时能使类内距离最小化,因此 DKL 技术的性能比 KL 变换提高了 10~30%。

2.3 离散余弦变换(DCT)

DCT 是由 DFT 导出的变换,对大多数自然图象而言,由于它的能量压缩性能接近于理想的 KL 变换,所以,目前几乎所有的图象、视频压缩标准,如 JPEG、MPEG 1、MPEG 2、H. 261/H. 263/H.26L 都使用了 DCT。由于上述标准在编码时都使用类似的算法,所以这里简单回顾一下基于 DCT 的 JPEG 标准算法^[9]。在 JPEG 编码中,编码分为计算 DCT、量化和变长编码 3 个步骤,即首先将原始图象分割为互不重叠的 8×8 的像素块;然后,对每个块进行二维离散余弦变换,并用 JPEG 建议的符合视觉特点的量化表进行量化,使每一个 8×8 的块生成 1 个 DC 系数和 63 个 AC 系数;接着对量化后的系数用“之”字形顺序重新排序,以形成一个一维的量化后系数序列;最后,对每一个像素块的 DC 系数用 DPCM 编码,而对剩余的 AC 系数则综合利用哈夫曼编码和游程长编码来进行压缩。标准 JPEG 编码器框图见图 4。显而易见,上述基于 DCT 的压缩标准中,并没有考虑图象检索方面的应用特点。

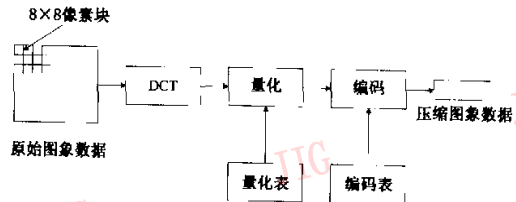


图 4 标准 JPEG 编码器

下面是近期出现的一些基于 DCT 的图象检索技术。

Smith 和 Chang 提出了一种将图象分割为 4×4 像素块的基于 DCT 的检索方法^[10]。该方法首先对每一个块计算出 16 个 DCT 系数,然后计算整幅图象的 16 个系数的均值和方差,再用得到的 32 维特征向量来代表图象的纹理特征。文献[10]用 Fisher 判别分析(FDA)来减少特征向量的维数。FDA 先从原始特征向量中生成一组线性参数,然后将这组精简后的特征值作为检索的索引。

Reeve 提出了一种类似于文献[10]的基于 DCT 的纹理识别方法^[11]。在这种方法中,图象先分割为 8×8 的像素块,并用前 8 个 AC 系数的方差构成特征向量。文献[11]假定这前 8 个 AC 系数是最具有区别能力的特征,从而避免了判别分析。由于这种方法生成的特征向量维数比上一种方法少,所以它的算法时间复杂度也比上一种方法小。

Shneier 和 Mottaleb 提出了一种对 JPEG 压缩图象进行检索的方法^[12]。这种方法是基于查询图象和目标图象的不相邻区域之间 DCT 系数的共有关系来进行检索。具体步骤是,首先在图象上选择 $2k$ 个窗口,并通过随机配对来形成 k 对窗口;然后对每一个窗口计算每一个 DCT 系数的均值,以生成 64 维的特征向量;最后通过比较每对窗口的特征向量,并根据它们的相似程度来对其赋值 0 或 1。这样,每对窗口就分配了 64bit,而查询图象和目标图象的相似程度就由所有的窗口对所分配的比特值决定。

Sim 和 Kim 提出了一种快速纹理描述和检索的方法^[13]。该方法首先对整个图象的 DCT 系数求均值,然后舍去高频系数,再对低频系数求取对数,以生成纹理描述符,而查询图象和目标图象的相似度则可用欧氏距离来测量。这种方法的突出特点是算法的时间复杂度很小,可运用到实时系统中。

Lay 和 Ling 提出了一种利用低频 DCT 系数的能量直方图进行检索的方法^[14]。文献[14]讨论了在选取不同范围的低频系数作为特征向量时检索效果的变化,比如,当只选取 DC 系数的能量直方图作为特征向量时,只有在查询图象和目标图象的颜色高度相似的情况下,才有较好的检索效果,而综合选取 DC 系数和若干低频 AC 系数的方法要优于只选取低频 AC 系数的方法。在选取 AC 系数个数的问题上,由于与 AC 系数对应的频率越高,量化效应对它的影响就越大,其带来的误差也就越大,所以,在块尺寸大小变化时,使用少数 AC 系数比使用较多 AC 系数要好,而且在整幅图象中有小的物体平移时,这

种直方图方法的检索结果也不受影响。

Yu 提出了一种计算 DCT 域两幅图象距离的方法^[15],即首先比较两幅图里每一个 8×8 像素块中的每一个 AC 系数,并根据相似程度将其赋值 0 或 1;然后对由 DC 系数形成的 DC 图象进行小波分解,再根据相似程度对两者产生的小波系数赋值 0 或 1;最后根据不同的权重计算出总的距离。这种方法的缺点之一就是查询图象的大小必须和目标图象一致。

许多基于内容的图象检索方法均是通过识别边缘信息来进行检索。Abdel-Malek 和 Hershey 提出了一种利用 DCT 系数来检测定向直线特征的方法^[16]。这种方法是基于以下观察:明显的水平、垂直、对角特征会使得 DCT 系数分别在水平、垂直、对角方向上的值较大。文献[16]指出,一条空域上斜率为 m 的直线在 DCT 域上可产生斜率近似为 $1/m$ 的直线。这种方法可推广到检测更为复杂的由直线组合成的特征。

Shen 和 Sethi 提出了一种从 JPEG 的 DCT 高频系数中检测出图象边缘和重要区域的方法^[17]。这种方法是从 8×8 像素块的 DCT 系数中估计出边缘方向、边缘相对中心的偏移以及边缘强度。其中,边缘方向包括水平、垂直、对角、基本水平和基本垂直 5 种方向。文献[7]实验结果表明,基于 DCT 的边缘检测有着和 Sobel 边缘检测算子相媲美的性能。

2.4 小波变换

近来,子带和离散小波变换(DWT)在图象编码和检索方面的应用日益流行^[18, 19]。在该方法中,图象递归地通过一系列低通、高通滤波器进行变换。在 DWT 中,低通输出进行递归滤波,而 Gabor 变换与小波变换类似。既然子带域的绝大部分能量都是用一部分低通系数表示的,那么舍弃高频系数将获得很高的压缩率,而子带编码则通常是通过正交镜像滤波器(QMFs)来减少由于数据抽取而带来的混叠效应。大家可以注意到,该方法中,图象数据是整幅通过一系列滤波器,而不像 JPEG 中那样分块。据研究,子带分解在编码方面有如下优点:(1)多分辨率;(2)对不稳定信号有适应能力;(3)具高度去相关和能量压缩性能;(4)可减少人工块效应和马赛克噪声;(5)更适应人类视觉系统的特点。

Chang 和 Kuo 提出了一种通过不规则树分解来进行纹理分析的方法^[20],在这种方法中,是采用中等分辨率的子带系数来做纹理匹配。由于其中 J

个最重要的子带能量组成了 J 维特征向量,因此检索时,只要匹配查询图象和数据库中目标图象的特征向量,即可实现检索.通过训练这种算法,即可得到每种纹理最重要的子带以及其平均能量,并可在纹理分类上达到较好的效果.

Smith 和 Chang 提出了一种纹理识别的方法^[10],该方法是首先利用子带能量来定义纹理特征序列,然后对图象进行 3 级离散小波变换,以生成 10 项特征向量,而纹理特征向量则通过 Fisher 判别来降低维数,然后用马氏(Mahalanobis)距离来计算查询图象和目标图象的相似度.

Chen 和 Kundu 利用小波变换和隐含马尔可夫模型(HMM)提出了一种对旋转和灰度级变换保持不变的纹理识别方法^[21]:第 1 步,从每一个子带中提取对灰度级变换保持不变的特征;第 2 步,子带序列用 HMM 来规范,并且每一类纹理有着自己的 HMM.研究表明,HMM 的主要用途在于它可以利用子带之间的相关性,而且可以获取旋转引起的变化趋势.其在进行纹理识别时,可将未知类别的纹理一一与所有模型匹配,以便从中找出最为匹配的模式作为未知类别纹理的模型.

Mandal 提出了一种通过比较方向性子带的直方图来进行纹理匹配的方法^[2].由于在摄像操作有所限定的情况下,相似的图象有着相似的子带直方图,因此借助不同级别的水平、垂直、对角信息的不同,可以区别不同图象.虽然不同图象在整个直方图上可能是相似的,但是不会具有类似的子带统计信息.实践证明,直接比较所有子带直方图的复杂度是很高的,而比较直方图的分布参数则可以大大降低复杂度.由于高通小波子带的直方图可以视为广义高斯密度函数分布^[22],因此其可用 σ (标准偏差)和 γ (形状参数)两个参数来表示.由此可见,通过比较子带直方图的这两个参数,就可以得到查询图象和目标图象之间的相似度.

绝大多数检索算法都假设图象的照度等级是相似的,但是,如果这个假设不成立的话,检索效果可能会大打折扣.Mandal 提出了一种小波域上基于直方图的图象检索方法^[23],这种方法对照度的变化有着较好的鲁棒性.由于它可通过直方图的等级不变矩来估测照度的变化,从而目标图象的子带参数 σ 和 γ 会随着照度的变化而做相应地调整.

Manjunath 和 Ma 提出了一种基于 Gabor 小波的图象检索方法^[24].该方法首先将每个图象分解为

4 个等级、6 个方向,然后用每个子带的均值 μ 和标准偏差 σ 构成一个 48 维的特征向量,而查询图象和目标图象间的相似度就由这个特征向量来计算.在这种方法中,由于小波域里的方向类别选择了 6 个,而不是 3 个(水平、垂直、对角),所以,这种方法有着较好的方向辨别性能.然而,由于 Gabor 小波在计算复杂度上比二进小波要高,所以这种方法的检索速度不高.

Wang 和 Chang 提出了一种子带域中,基于相关性的模式匹配法^[25].文献[25]根据两个空域函数的子带系数导出了它们的相关性表达式,即考虑两个一维信号 $x(n)$ 和 $w(n)$,这两个信号的相关性可以用 Z 变换域中的 $W^*(z)X(z)$ 来表示;设 $H_0(z)$ 为低通滤波器, $H_1(z)$ 为高通滤波器,假设 $x(n)$ 和 $w(n)$ 经过子带分解后分别得到 $\{y_0(n), y_1(n)\}$ 、 $\{z_0(n), z_1(n)\}$,则它们之间的相关性如下

$$W^*(z)X(z) = \sum_{i,j=0}^1 F_{i,j}(z)Z_i^*(z^2)Y_j(z^2)$$

其中, $F_{i,j}(z) = H_i(z)H_j^*(z)$, $i, j = 1, 0$. 上角 * 表示共轭.

由上式可知,两个信号的相关性等于它们在子带域相关性的加权和.在一些具有高能量的子带上进行计算(这样可以大大降低计算量),不仅可以得到对相关系数峰值的合理估计^[25],而且,由于这些系数同时也提供了空间信息,因此对子带也不需要再进行合成分析.而在傅立叶变换域中,如要求得相关系数峰值,则必须进行反变换.

Jacob 提出了一种直接比较 DWT 系数的检索方法^[26].这种方法是首先对所有图象进行小波变换,并使其大小都规范到 128×128 pixels;然后将计算得到的 M 个(作者采用 40~60)最大幅度的小波系数的平均颜色、符号(正号或负号)作为索引;再将数据库中所有图象的索引规范成单一的数据结构,以便于快速图象检索.实践证明,这种方法虽能提供较好的索引功能,但这样生成的索引需依赖于 DWT 系数的位置,如果目标图象是由查询图象经平移或旋转之后形成的,那么在检索时就可能会被漏掉.

Wang 提出了一种与 Jacob 的方法类似的技术^[27].这种方法首先对所有图象进行 4 级小波分解,并使其大小都规范到 128×128 pixels,同时将分解后的 4 个 8×8 的最低分辨率图象分别用 S_L (低通)、 S_H (水平频带)、 S_V (垂直频带)、 S_D (对角频带)

来表示;然后,分以下3个步骤来完成图象匹配:(1)对20%的图象用 S_L 频带的方差来进行检索;(2)根据查询图象和目标图象的 S_L 系数的差值,一部分图象将被选出来;(3)根据查询图象和目标图象的 S_L 、 S_H 、 S_V 、 S_D 系数的差值检索出最终结果.对于颜色图象,在颜色通道上重复整个过程.这种方法在计算复杂度上比分级匹配的方法要小,与Jacob的方法^[26]相比,也有所改进,但是,它仍然对平移和旋转不能保持鲁棒性.

Qi提出了一种复数小波变换方法^[28],在这种变换中,DWT系数的幅度是旋转不变的,其母小波在极坐标下被定义.通过对一系列英文字母图象进行的实验结果证明,这种方法优于复数Zernike矩(它的幅度同样是旋转不变的).Rashkovskiy提出了一类对灰度级、图象旋转、移动(SRS)变化保持不变的非线性小波变换方法^[29].这种小波变换可根据输入信号的不同来调整母小波,以保持对SRS变化的不变性.在每一次实际应用时,它的小波系数或者说小波形状经过迭代计算,能使得能量函数最小化.

Froment和Mallat提出了一种第2代图象编码方案^[30],这种方案能从纹理信息中分离出边缘来.该方法首先根据小波变换系数的局部模极大值来检测出多尺度边缘;然后,用原始图象减去重构后的图象来得到提供纹理信息的误差图象;最后用标准正交小波变换对纹理进行编码.实验结果表明,这种多尺度边缘信息有着较好的图象检索能力.

2.5 矢量量化

矢量量化是一种非常适合于极低比特率环境的编码方法.它的主要思想是通过训练集合首先构造一个有限矢量个数的码书,然后对任意输入图象中的每一分块矢量按一定的距离准则进行分类,并且将原图象中的所有分块矢量都采用与之相匹配的特征矢量码字来进行表征的,而与所有分块矢量相匹配码字的索引序列就是最终的压缩数据.这样的编码结构和压缩数据非常适合与图象检索技术相结合,其主要体现在:矢量量化本质上是一个聚类和分类的过程,这样的结构天然地和图象聚类分析以及分类处理等操作的要求相一致;由于矢量量化是通过一个有限数量的再生矢量集合来对所有输入图象进行表征的,所以对该矢量集合(码书)的任何分析处理操作都可以间接地反映到解压缩图象中.正是因为矢量量化所具有的这种良好特性,所以使其能广泛应用于图象增强、图象分类、边缘检测以及图象

检索等方面^[31].

矢量量化器(VQ)如同一个映射 Q ^[32],这个映射将 K 维欧氏空间映射到一个有限子集 Y ,如 $Q: R^{<K>} \rightarrow Y$,其中, $Y = \{\hat{x}_i; i=1, 2, \dots, N\}$ 是一组再生矢量,被称为VQ码书或VQ表, N 是 Y 中矢量的个数.VQ编码映射如图5所示^[32].根据距离最近原则,VQ编码器将每一个输入的矢量映射到一组有限码字(码书)上,而码字的标签(索引)则用来表示输入图象.

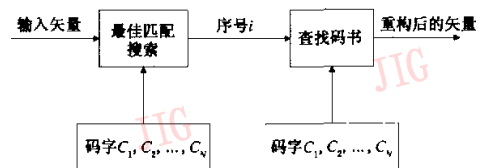


图5 矢量量化框图

由于在矢量量化过程中,任一输入图象的分块矢量都依据统一的距离准则映射到有限特征矢量集合(公共码书)的某一码字上,所以不同的图象与该有限矢量集合之间映射的统计特性在一定程度上反映了图象的内容. Idris和Panchanathan基于这样的思想,构造了图象特征矢量集合中的矢量使用频数统计直方图,并将其作为表征图象内容的特征量,由于可将其用于比较查询图象和图象库中的目标图象的相似度,从而实现了将矢量量化压缩数据直接用于支持图象检索的目的^[33~37].同时,为了进一步降低对直方图特征量的存储要求以及直方图相互间距离计算的复杂度,他们还还对码字使用频数直方图进行了简化处理,提出了使用图(usage map)特征量来对码字进行简化处理^[36]的方法,即某一码字矢量,如果在图象压缩过程中使用过,则与该码字对应的特征位置为1,否则为0,而不同图象间相似程度的计算只需对其使用不同的图特征量按位进行异或操作(求汉明距),即可大大减小特征向量的存储量和距离计算的复杂度,但这样简化处理的代价是降低检索的正确率.基于统计特征量的矢量量化域的图象检索方法的主要不足在于,由于统计直方图特征量无法反映码字使用的位置信息,从而使其检索效率难以提高.同时,由于基于统计特征量的检索结果无法预测,因此也导致实际中的检索结果不易满足用户的需要^[37].

Barbas和Wolk探讨了大型雷达图象数据库的有效表征问题^[36],目的是为了优化存储量和搜索时

间,他们利用树状结构矢量量化的聚类方式,并用树状结构矢量量化下的多分辨率小波来表征图象特征,这种树状结构是根据脉冲的多分辨率小波分解结果引入的,因此该方法的整体性能较好。

Vellaika 将 VQ 技术应用到基于内容的遥感图象检索中^[39]。其在应用中,是采用估测多种形变的方式来提高作为内容描述符(content descriptor)VQ 码字的性能。文献^[39]通过测试两种查询(按类查询、按数值查询)的性能,发现后者的性能更为卓越。

2.6 分形/仿射变换

自从 Jacquin 首先提出迭代函数系统(Iterated Function Systems, IFS)的正式数学描述以来, Barnsley 又将迭代函数系统引入到计算机图形学应用中,并提出分形图象压缩概念,至今分形编码一直是一种倍受注意的编码方式,许多学者在这方面都进行了不少的研究。大家知道,分形编码是建立在迭代函数系统和拼贴原理(Collage Theorem)等数学理论基础上的。在分形编码中,有排列块(Range)和区域块(Domain)两个重要元素。其中,排列块虽然是互不重叠的,但区域块却有可能互相重叠,由于区域块的尺度比排列块要大,而且在一般情况下,区域块的边长是排列块边长的两倍,因此在分形编码时,编码器就利用每个给定的排列块来搜索与该排列块最相似的区域块,其仿真变换参数就是常说的分形码。大家知道,这种排列块和区域块之间的相似变换 W 是由几何变换和亮度变换两部分组成,而且据分形编码原理,变换 W 必须是收敛的。这样,图象就可以用一组分形码表示,这就消除了图象数据的冗余信息。根据迭代函数系统和拼贴原理,原始图象可以由这些分形码通过多次迭代得到很好的近似。由此可看出,由于分形编码是基于图象局部区域间的相似性来进行编码的,而且,分形码还能够一定程度上反映图象中各目标间的空间位置关系,也有助于对图象内容的描述,所以,用分形编码技术不仅可以避免抽取可视化特征的困难,而且又可以对图象内容进行描述^[40]。

Sloan 首先发现了分形图象编码技术在基于内容的图象检索中的应用前景,并进行了探索^[41]。他提出了将图标图象数据库中的图象直接拼凑起来进行编码的方法,即对于任何图象中的排列块可以在任何图象中找到区域块,而两幅图象之间的相似性则可以用一个图象中的子块当作另一个图象中区域块的数目来度量。对这方面的进一步研究是由

Zhang 等人开展的,他们分别提出了基于联合分形编码(Joint Fractal Coding)的图象检索策略^[43]以及基于九分树分解和分形编码的图象检索策略^[42]。尽管许多学者在提高分形编码时间上作了很大的努力^[44~46],但是总的说来,分形编码仍然是一种非常耗时的编码方法。

2.7 混合法

在图象、视频压缩中,混合编码法通常指的是结合两种以上基本编码方法的编码方法^[19,47]。这些混合编码法利用了联合的多种压缩技术的优点,从而能提供更好的编码性能。

Idris 和 Panchanathan、Swanson 都提出了基于小波域的矢量量化的图象检索方法^[34,48]。Idris 和 Panchanathan 通过对小波系数进行矢量量化来形成索引,而 Swanson 则使用小波和 VQ 联合的方法来对图象进行检索。

2.8 各种压缩域的比较

对多种压缩域方法的图象检索性能进行比较是困难的,因为,一种图象检索方法成功与否、流行与否,需依赖于与之相关的编码方法性能。前面在分类介绍各种压缩域的检索方法时,已经分析过它们各自的优缺点,下面再从压缩变换本身进行一番比较。

KL 变换虽然在统计上是最优的,但是计算复杂度庞大;而且必须存储基本图象,从而导致较差的编码性能。JPEG 的分块离散余弦变换虽有着良好的编码和检索性能,但由于分块结构并不是为图象检索而构造的,因此在检索的具体应用中,可能会造成一些问题。由于基于小波变换的方法,由于具有小波内在的多分辨率、易于检测边缘和形状、易于检测方向信息等优点,因此检索性能较为可靠。从文献^[10]中可知,在纹理分类上,小波变换比 DCT 变换性能更好,因为 VQ 使用索引进行编码,所以它有着天生的图象检索能力,即利用相同的索引就可以进行图象检索。分形虽然有着良好的编码能力,但它是高度非线性的,同时是依赖于图象的,而且,直接利用分形码进行检索的话,检索的正确率不高。文献^[41]对分形和小波变换进行了比较,得出的结论是:在对包含明显纹理特征的图象进行检索时,小波变换的效果较好;在对多种类型图象进行检索时,分形变换的整体效果较好。另外,从检索算法的计算量方面考虑,VQ 和分形编码的复杂度都是高度非对称的,例如,编码的计算量很大,而解码则相反,这使得基于分形和 VQ 的检索效率难以提高。

3 压缩域图象检索的具体应用

Bae 和 Jung 提出了一种利用 DCT 系数进行纹理检索的系统^[49]。该系统由输入模块、查询模块和检索模块 3 个主要部分组成。在输入模块中,一方面特征向量从输入图象中抽取,并和输入图象一起存储在图象库中;另一方面,当查询图象输入到查询模块时,即可首先抽取它的特征向量;然后,在检索模块中,与图象库中存储的特征向量比较,再根据匹配程度检索出相似的图象;最后,即可从检索出的多幅图象中获得目标图象。

文献^[49]利用压缩域的部分 DCT 系数作为特征向量来进行检索。由于 DCT 系数有以下特点:(1)每个子块的 DC 系数代表了图象平均能量的部分;(2)子块中剩下的 DCT 系数包含了造成图象差异的频率信息;(3)在子块中某些区域的系数代表了某些方向信息(例如,DCT 变换域最上方和最左方的系数分别代表了部分垂直和水平边缘信息),这样从每一子块的 64 个 DCT 系数中可选出 9 个特征向量。在相似度测量部分,为了在大型图象数据库中减少搜索空间,以便更有效地进行检索,作者使用分层相似度测量法,即,首先进行全局搜索,并使用纹理类别码字(由 ART2 神经网络^[50]生成)来进行粗略的相似度测量,同时将与查询图象具有相同类别码字的图象挑选出来;然后,在本类别的图象之中作局部搜索,同时使用某个特征向量来进行精确的相似度测量,以得到最终的检索结果;最后,用修正后的欧氏距离 $D(H, I)$ 来衡量图象 H 和图象 I 的相似度

$$D(H, I) = \sum_{s=1}^k \sum_{r=1}^n |E_{s,r}^{(H)} - E_{s,r}^{(I)}|$$

其中, $E_{s,r}^{(H)}$ 是图象 H 的 s 行, r 列的一个特征向量, $E_{s,r}^{(I)}$ 是图象 I 的 s 行 r 列的一个特征向量。

这种方法与 Tamura 的方法^[51]比较,查全率提高了 42%,准确率提高了 64%,速度快 4 倍。平均检索率超过了 82%。

4 发展前景

从上述讨论可以看出,压缩域图象检索技术是可行,而且是有效的。另外,压缩和检索相结合的方式很适合新一代图象处理应用的要求,具有巨大的应用潜力。尽管如此,作为一个新兴的研究方向,压缩域的图

象检索技术才刚刚起步,还有很多问题需要深入研究:一方面,压缩域的图象检索作为一种基于内容的检索技术,面临许多和传统检索技术类似的问题;另一方面,作为一种压缩域上的图象处理技术,它又有着与一般压缩域处理技术相似的发展前景。

同传统的图象检索技术一样,压缩域图象检索是一种综合集成技术,其中的一些关键技术,如特征提取、图象分割等等是图形图象界的经典难题。其困难本质是,在本身不含任何语义信息的原始图象数据上进行基于内容的检索,现存的各种理论和技术都不甚完善,比如,如何综合多种特征,如何利用特征之间的互补能力来提高检索效率;对于高维索引,还需研究支持快速检索的索引结构和算法;还需建立充分反映图象内容的数据模型等等。

作为一种压缩域上的图象处理技术,在理论上,需要对图象检索技术在原始数据域和压缩数据域所表现出来的不同形式进行研究,以便更好地实现压缩域的检索,同时,还需要研究能全面客观评价压缩域图象检索技术的方法;在技术上,由于对基于 DCT 压缩域的图象检索技术的研究已经进行了多年,而随着 JPEG 2000 以及 MPEG-4 标准的制定,将有越来越多的图象、视频信息采用基于小波变换的压缩算法进行压缩,所以,基于小波域的图象处理技术的研究将来会成为一个热点^[52]。另外,随着社会信息化和网络化程度的进一步加深,操作环境多元化的问题日益突出,信息社会的最终目标是能够使任何人在任何时刻、任何地点都能通过适当的手段获取到他所需要的信息或向他人提供信息,这就相应带来了操作环境多元化的问题。首先是网络环境的多元化,如可以是宽带的网络,如局域网、Cable Modem 或是 XDSL,也可能是窄带的 PSTN、无线信道等;其次是终端设备的多元化,根据环境的不同,用户所使用的可能是计算机、智能家电,也可能是比较简单的手持设备,因为不同终端设备对应不同的接受处理能力,此外,还有媒体的多元化以及用户的多元化等。总之,新一代图象所具有的海量性、无序性以及操作环境多元化等特点,已向新的图象应用和业务的开发提出了新的挑战。面对各种各样的多元化,新的压缩域图象检索技术的研究必须使其具备自适应的能力,即能够根据操作环境的变化提供相应的服务,以满足多元化操作环境的需要。

从长远来看,支持压缩域图象检索的压缩算法的研究任重道远,如何通过设计实现新一代的压缩

算法来有效地支持压缩码流的处理、分析的功能,将成为压缩领域研究的重点^[53]。在支持检索的图象压缩算法中,由于分形法是在图象空间和迭代函数系统空间建立一一对应的映射关系,且用分形编码技术既可以避免抽取视觉特征的困难,又可有效地表征图象内容,从而具有相当潜力^[54]。

开展压缩域图象信息处理技术的研究,就是把图象信息的压缩算法和它所能支持的基于内容的处理技术综合起来进行系统而深入的研究。这一方面是为了使得对图象的处理操作能够直接在压缩码流上进行;另一方面是为了设计并实现新的图象压缩算法,使其不仅具有较高的压缩效率和信噪比,还能支持图象信息基于内容的分析、处理和检索等功能,并且使压缩码流具有精细的可分级性。另外,还可以进一步建立基于 WWW 的对图象压缩数据进行分析、处理和检索的实验系统,用于验证上述研究所提出的技术和算法,并为构架信息网上的新业务和新应用提供技术基础和原型系统。这样就不仅要从数据压缩技术的角度来进行压缩域处理技术的研究,而且还要围绕矢量量化、分形和小波等编码压缩算法及它们所能支持的图象检索、分析处理等功能展开研究^[1]。

参 考 文 献

- 1 沈兰荪. 压缩域图像/视频信息处理技术的研究[J]. 计算机自动测量与控制, 2000, 8(5): 1~3.
- 2 Mandal M K, Aboulnasr T, Panchanathan S. Image indexing using moments and wavelets [J], IEEE Transactions on Consumer Electronics, 1996, 42(3): 557~565.
- 3 Mandal M K, Idris F, Panchanathan S. A critical evaluation of image and video indexing techniques in the compressed domain [J]. Image and Vision Computing, 1999, 17(7): 513~529.
- 4 Stone H S, Li C S. Image matching by means of intensity and texture matching in the Fourier domain[A]. In: Proc. of SPIE: Storage and Retrieval for Image and Video Databases IV[C], San Jose, CA USA, 1996, 2670: 337~349.
- 5 Augusteijn M, Clemens L E, Shaw K A. Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier[J], IEEE Transactions on Geoscience Remote Sensing, 1995, 33(3): 616~626.
- 6 Celentano A, Lecce V D. A FFT based technique for image signature generation [A]. In: Proc. of SPIE: Storage and Retrieval for Image and Video Databases V[C], San Jose, CA USA, 1997, 3022: 457~466.
- 7 Pentland A, Picard R W, Sclaroff S. Photobook: Tools for content-based manipulation of image databases[A]. In: Proc. of SPIE: Storage and Retrieval for Image and Video Databases I [C], San Jose, CA USA, 1994, 2185: 34~47.
- 8 Swets D L, Weng J. Using discriminant eigenfeatures for image retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 831~836.
- 9 Wallace G K. The JPEG still picture compression standard[J]. Communication of the ACM, 1991, 34(4): 31~45.
- 10 Smith J R, Chang S F. Transform features for texture classification and discrimination in large image databases [A]. In: Proc. of IEEE International Conference. Image Processing [C], Austin, TX USA, 1994, 3: 407~411.
- 11 Reeves R, Kubik K, Osberger W. Texture characterization of compressed aerial images using DCT coefficients[A]. In: Proc. of SPIE: Storage and Retrieval for Image and Video Databases V [C], San Jose, CA USA, 1997, 3022: 398~407.
- 12 Shneier M, Mottaleb M A. Exploiting the JPEG compression scheme for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 849~853.
- 13 Sim D G, Kim H K, Park R H. Fast texture description and retrieval of DCT-based compressed images [J], Electronic Letters, 2001, 37(1): 18~19.
- 14 Lay Jose A, Ling Guang. Image retrieval based on energy histograms of the low frequency DCT coefficients[A]. In: Proc. of International Conference on Acoustics, Speech and Signal Processing [C], Phoenix, Arizona, USA, 1999, 6: 3009~3012.
- 15 Yu Hong Heather. Visual image retrieval on compressed domain with Q-distance [A]. In: Proceedings. Third International Conference on Computational Intelligence and Multimedia Applications ICCIMA'99. [C], New Delhi, USA, India, 1999, 285~289.
- 16 Abdel-Malek A A, Hershey J E. Feature cueing in the discrete cosine domain[J]. Journal of Electronic Imaging, 1994, 3(1): 71~80.
- 17 Shen B, Sethi I K. Direct feature extraction from compressed images[A]. In: Proc. of SPIE: Storage and Retrieval for Image Video Databases IV[C], San Jose, CA USA, 1996, 2670: 404~414.
- 18 Mallat S G. A theory for multiresolution signal representation: the wavelet decomposition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 674~693.
- 19 Antonini M, Barlaud M, Mathieu P *et al.* Image coding using wavelet transform[J], IEEE Transactions on Image Processing, 1992, 1(2): 205~220.
- 20 Chang T, Kuo C C J. Texture analysis and classification with tree-structured wavelet transform [J]. IEEE Transactions on Image Processing, 1993, 2(4): 429~441.
- 21 Chen J L, Kundu A. Rotation and gray scale invariant texture identification using wavelet decomposition and hidden Markov model[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(2): 208~214.
- 22 Birney K A, Fischer T R. On the modeling of DCT and subband image data for compression [J]. IEEE Transactions on Image Processing, 1995, 4(2): 186~193.
- 23 Mandal M K, Panchanathan S, Aboulnasr T. Image indexing using translation and scale-invariant moments and wavelets[A]. In: Proc. of SPIE: Storage and Retrieval for Image and Video Databases V[C], San Jose, CA USA, 1997, 3022: 380~389.
- 24 Manjunath B S, Ma W Y. Texture features for browsing and retrieval of image data[J]. IEEE Transactions. Pattern Analysis and Machine Intelligence, 1996, 18(8): 837~841.
- 25 Wang H, Chang S F. Adaptive image matching in the subband domain[A]. In: Proc. of SPIE, VCIP[C], San Jose, CA USA,

- 1996, 2727; 885~896.
- 26 Jacobs C E, Finkelstein A, Salesin D H. Fast multiresolution image querying[A]. In: Proc. of ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques [C], Los Angeles, CA, USA, 1995: 277~286.
- 27 Wang J Z, Wiederhold G, Firschein O *et al.* Wavelet-based image indexing techniques with partial sketch retrieval capability [A]. In: Proc. of the Forum on Research and Technology Advances in Digital Libraries [C], Washington, DC, USA, 1997, May; 13~24.
- 28 Qi F, Shen D, Quan L. Wavelet transform based rotation invariant feature extraction in object recognition[A]. In: Proc. of Int. Symp. on Information Theory & its Applications [C], Sydney, Australia, 1994, November; 221~224.
- 29 Rashkovskiy P, Sadovnik L. Scale, Rotation and shift invariant wavelet transform [A]. In: Proc. of SPIE; Optical Pattern Recognition [C], San Jose, CA USA, 1994, 2237; 390~401.
- 30 Froment J, Mallat S. Second generation compact image coding with wavelets [A]. In: C. K. Chui (Ed.), Wavelets; A Tutorial in Theory and Applications [M], New York; Academic Press, 1992.
- 31 沈兰荪, 魏海, 黄祥林. 压缩域图像处理技术研究[J]. 北京工业大学学报, 2000, 26(3): 24~32.
- 32 Gersho A, Gray R M. Vector quantization and signal compression [M]. Boston, MA USA; Kluwer Academic Publishers, 1991.
- 33 Idris F, Panchanathan S. Image indexing using vector quantization [A]. In: Proc. of SPIE; Storage and Retrieval for Image and Video Databases III [C], San Jose, CA USA, 1995, 2420; 373~380.
- 34 Idris F, Panchanathan S. Storage and retrieval of compressed images [J]. IEEE Transactions on Consumer Electronics, 1995, 41(3): 937~941.
- 35 Idris F, Panchanathan S. Image indexing using wavelet vector quantization [A]. In: Proc. of SPIE; Digital Image Storage Archiving System [C], San Jose, CA USA, 1995, 2606; 269~275.
- 36 Lee J, Dickinson B W. Multiresolution video indexing for sub-band coded video database [A]. In: Proc. of SPIE; Storage and Retrieval for Image and Video Databases III [C], San Jose, CA USA, 1995, 2420; 162~173.
- 37 Jiang J. An image compression and indexing system using neural networks [J]. Journal of Visual Communication and Image Representation, 1997, 8(2): 135~145.
- 38 Barbas J S, Wolk S I. Efficient organization of large ship radar databases using wavelets and structured vector quantization [A]. In: Proc. of Asilomar Conference on Signals, Systems and Computers [C], Pacific Grove, CA USA, 1993, 1; 491~498.
- 39 Vellaikal A, Kuo C C J, Dao S. Content-based retrieval of remotesensed images using vector quantization [A]. In: Proc. of SPIE; Visual Information Processing IV [C], San Jose, CA USA, 1995, 2488; 178~189.
- 40 王志勇, 池哲儒, 余英林. 分形编码在图像检索中的应用[J]. 电子学报, 2000, 28(6): 19~23.
- 41 Sloan A D. Retrieving database contents by image recognition; New fractal power [J]. Advanced Imaging, 1994, 9(5): 5.
- 42 Zhang Aidong, Cheng Diao, Acharya Raj. An approach to query by texture in image database systems [A]. In: Proc. of SPIE; Digital Image Storage Archiving System [C], San Jose, CA USA, 1995, 2606; 338~349.
- 43 Zhang Aidong, Cheng Diao, Acharya Raj. Menon raghu, comparison of wavelet transform and fractal coding in texture based image retrieval [A]. In: Proc. of SPIE; Visual Data Exploration and Analysis III [C], San Jose, CA USA, 1996, 2656; 116~125.
- 44 张檀, 陈刚, 金以文. 基于金字塔正交小波分解的快速分形图像编码[J]. 电子学报, 1998, 26(8): 37~42.
- 45 Jacquin A E. Fractal image coding; A review [J]. In: Proc. of IEEE, 1993, 81(10); 1451~1465.
- 46 Fisher Y. Fractal image compression: Theory and application [M]. New York; Springer Verlag, 1995.
- 47 Li J. Hybrid wavelet-fractal image compression based on a ratedistortion criterion [A]. In: Proc. of SPIE; Visual Communications and Image Processing [C], San Jose, CA USA, 1997, 3024; 1014~1025.
- 48 Swanson M D, Hosur S, Tewfik A H. Image coding for content-based retrieval [A]. In: Proc. of SPIE; Visual Communications and Image Processing [C], San Jose, CA USA, 1996, 2727; 4~15.
- 49 Bae Hee-Jung, Jung Sung-Hwan. Image retrieval using texture based on DCT [A]. In: Proceedings of the International Conference on Information, Communications and Signal Processing, ICICSP [C], Singapore, 1997, 2; 1065~1068.
- 50 Carpenter Gail A, Grossberg Stephen. ART 2; self-organization of stable category recognition code for analog input pattern [J]. Applied Optics, 1987, 26(23); 4919~4930.
- 51 Tamura H, Mori S, Yamawaki T. Textures corresponding to visual perception [J]. IEEE Transactions on Systems, Man and Cybernetics, 1976, 8(6); 460~473.
- 52 Wolf L C. Multimedia communication [J]. In: Proc. of IEEE, 1997, 85(12); 1915~1933.
- 53 Chang S F. Compressed-domain technology for image/video indexing and manipulation [A]. In: IEEE International Conference Image Processing '95 on Special Issue on Intelligent Signal Processing, Washington DC, USA, 1995; 1~11.
- 54 Wei H, Shen L. Fractal-based image storage and indexing [A]. In: Proc. of SPIE; Storage and Retrieval for Media Databases [C], San Jose, CA USA, 2000, 3972; 421~429.



黄翔宇 1978年生, 2000年毕业于清华大学电子工程系, 现在清华大学电子工程系攻读硕士学位. 主要研究方向为图象检索.

张乾 教授, 博士生导师. 研究领域主要是由其倡导的图象工程(图象处理、图象分析、图象理解及其应用)学科. 已发表180余篇研究论文, 著有《图象分割》等书6本. 现为IEEE高级会员.