

基于视觉特性的静态图象压缩编码

沃焱 韩国强

(华南理工大学计算机科学与工程学院, 广州 510641)

张波

(广东省中山市公安局, 中山 528400)

摘要 基于主观视觉特性的压缩编码方法突破了传统编码技术中以MSE作为评判标准的限制, 将人眼视觉特性结合到编码算法中, 以实现最低比特率和最好视觉效果。目前该技术的研究已有较大进展, 但仍存在许多问题有待研究解决。本文系统地介绍了现有的基于视觉特性的压缩编码方法, 同时讨论了应用在JPEG-2000和MPEG-4中的一些视觉编码技术。通过研究总结出以下结论: 图象质量的客观评价、感兴趣区域的自动确定以及任意对象的识别与分割是基于视觉特性的编码技术中存在的主要问题, 也是今后进一步研究发展的方向。

关键词 计算机图象处理(520·6040) 视觉特性 静态图象压缩 刚能鉴别的视觉失真(JND) 视觉感知

中图法分类号: TN919.8 **文献标识码:** A **文章编号:** 1006-8961(2003)10-1105-10

Still Image Coding Based on Models of Visual Characteristic: An Overview

WO Yan, HAN Guo-qiang

(College of Computer Science and Engineering, South China University of Technology, Guangzhou 510641)

ZHANG Bo

(Zhongshan police office, Guangdong 528400)

Abstract The visual-characteristic-based compression technique breaks through the restriction of the traditional image compression technique that uses MSE as criterion, and combines the human visual characteristic with the coding arithmetic, in order to achieve the lowest bit rate and the best perceptual image quality. Now, much headway has been made in the research of this technique, but there still are many problems to be researched on. This paper systematically introduces the existing methods of this compression technique, and discusses some visual coding techniques that applied in JPEG-2000 and MPEG-4 standards. After researching on the techniques, Some results and conclusion are given as follow: perceptual measurement of the image quality, automatic ascertainment of the region-of-interest and recognition and segmentation of the arbitrary object are the main problems existing in the visual-characteristic-based coding technique, and they are also the research direction for the future.

Keywords Visual characteristic, Still image coding, Just noticeable distortion, Visual perception

0 引言

数字图象含有非常大的数据量, 它与通信网容量的矛盾及其传输和存储的困难都极大地制约了数字图象通信的发展, 并一度成为数字图象通信发展中的瓶颈问题。图象压缩编码是各类图象信息传输系统、多媒体系统和存储系统的一项十分关键的技术。图象压缩编码最根本的目的就是要以尽量少的比特数来表征图象, 同时要保持恢复图象的质量, 使

之符合特定应用场合的要求。长期以来, 传统的压缩编码方法是基于最小均方误差(MSE)准则的编码方法。均方误差从总体上反映原始图象和恢复图象的差别, 因对图象中所有像点同等对待, 故其并不能反映局部和人眼的视觉特性, 用MSE作为恢复图象质量的评判标准所求得的结果常与人们主观视觉效果不一致。对于图象信号, 人眼是最终的信号接收者, 因此在压缩编码时应当考虑人的主观视觉效果, 而不仅以MSE作为评判标准。由于人的视觉能力有限, 对有些失真很难觉察, 因此图象经过压缩编码

基金项目: 国家自然科学基金(10171033); 留学回国人员科研基金; 广东省“千百十”优秀人才基金

收稿日期: 2003-01-30; 改回日期: 2003-04-21

后允许恢复图象存在一定的失真,但失真要在人眼视觉可容忍的范围之内或满足特定的应用要求。如何将人眼视觉特性结合到压缩编码技术中,在提高压缩比的同时又不损害图象的恢复效果?在低比特编码时,如何按照人眼视觉特性来优化比特分配,使得恢复图象具有最符合人眼视觉效果的主观质量?这些问题一直是图象压缩编码中的一个热点问题。近年来人们采用基于主观视觉特性的压缩编码方法,取得了很大的进展,其已被应用到电视编码传输^[1]、高清晰度电视编码传输^[2]以及 JPEG^[3]、JPEG 2000^[4,5]、MPEG-4^[6]压缩标准中,但该技术还有许多问题有待于进一步解决。与传统的以去除像素间的数据冗余为目标的编码方法不同,基于主观视觉特性的压缩编码方法在编码时将人眼的视觉特性结合到编码算法中,以设计最低比特编码,最好视觉效果为目的。基于人眼视觉特性的压缩编码方法主要分为3类:一是基于感知的压缩编码方法;二是基于人眼感兴趣区域(ROI)的压缩编码方法;三是基于轮廓边缘和纹理的压缩编码方法。

1 人眼视觉特性

长期以来,研究者们通过对人眼视觉现象的观察与研究,发现人类的视觉有许多特点,在图象编码中如能充分利用这些视觉特性来去除图象中的视觉冗余,可在保持主观质量不变的条件下降低比特率、提高压缩比。

1.1 Weber 定律

人眼主观上刚刚可辨别亮度差别所需要的最小光强差值 ΔI 称为亮度的辨别阈值^[7]。只有当刺激光强度 I 增加到某一值之后,才能感觉到其亮度变化。若背景宽广且亮度均匀,则 ΔI 在很大范围内近似与 I 成比例: $\Delta I/I \approx 0.02$ 。但在 I 非常高或者非常低时, $\Delta I/I$ 则不是常数,一般称 $\Delta I/I$ 为视觉的对比灵敏度。

1.2 视觉的空间和时间频率特性

当快速运动的物体从眼前通过时,人眼很难看清其细节,只能看出其粗略轮廓。只有当物体细节大小、细节明暗对比度以及在眼中呈现的时间长短都比较合适时,才能对物体的细节有清楚的识别。这种特性可以用视觉空间-时间频率特性来定量描述。对亮度在空间中作正弦变化的光栅,通过求出人眼对各种节距正弦显示的图案的认识程度,可以得到视

觉空间频率特性。如果在视野内将亮度作正弦变化,则得出视觉的时间频率特性^[8]。视觉系统的空间频率特性和时间频率特性是相互依赖的,如图1所示。

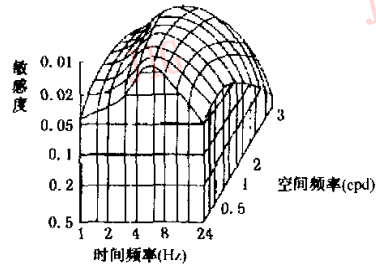


图1 人眼视觉系统的时-空灵敏度

1.3 视觉的阈值效应和掩盖效应

视觉阈值是指正好可以被看到的刺激(干扰或失真)值。在亮度有变化的边缘上,该边界“掩盖”了边缘邻近像素的信号感觉,使人的感觉变得不灵敏、不精确,这种效应称为掩盖效应。边缘的掩盖效应与眼睛的无意识运动、边缘出现的时间长短以及运动状况有关。视觉阈值随图象内容的变化而变化,在平坦区域阈值低,对失真也较敏感;而在边缘区和纹理区,视觉存在掩盖效应,阈值对失真不敏感。视觉阈值效应和掩盖效应被广泛用来设计图象量化编码器,以提高压缩比,即利用视觉的这种特性在图象的边缘区域可以容忍较大的量化误差,从而减少量化级,并降低数码率。影响视觉阈值的因素有两个:一是对比灵敏度,它说明了阈值取决于干扰和或失真所处的背景的亮度(总效应);二是某点的亮度阈值的掩盖效应,它与邻近像素的邻域干扰或失真有关(局部效应)。

1.4 眼球的运动及控制

由中央凹中锥状细胞构成的分辨力高的中心视力一般小于 10° ,而视网膜周围由杆状细胞构成的周边视力分辨力低,不能看清图象细节,但周边视力对图象中运动变化部分很敏感,有特征抽取作用从而控制眼肌转动视轴,使中心视力对准这些部位,以便看清其细节。但是,光靠这种转动无法形成整个画面的印象,必须依靠周边视力。

在对视觉特性研究的基础上,人们将视觉特性应用到压缩编码中。根据人眼的视觉阈值效应和掩盖效应,提出了基于视觉感知的编码方法^[16~28]。根据人眼在观察景物时需要转动眼球,将注意点分布在特征突出的部位这一视觉特性,提出了基于人眼

感兴趣区域(ROI)的编码方法^[29~46];综合人眼对轮廓边缘失真比较敏感的特性,提出了基于轮廓边缘和纹理的编码方法^[41~53]。

2 基于视觉感知的编码

由于人眼具有视觉阈值效应和掩盖效应(统称为感知特性),因此人的视觉分辨率有限,可以容忍一定的图象失真。研究表明,频率、背景亮度、纹理都是影响人眼辨别景物失真的因素^[8],因此需综合考虑这些影响因素,根据人眼在不同空间频率、不同内容的图象数据中的失真敏感度来确定比特分配。对于人眼失真敏感的区域编码时要多分配些比特进行精细量化,对于人眼失真不太敏感的区域可以少分配些比特进行粗糙一些的量化,从而达到最好视觉感知效果的最优比特分配。人眼对不同区域失真的敏感性可以用刚能鉴别的视觉失真(Just Noticeable Distortion 简称为 JND) 阈值来度量。Jayant 认为基于视觉感知的编码方法实质上是基于刚能鉴别的视觉失真(JND)阈值的编码方法^[8]。所谓刚能鉴别的视觉失真阈值,即当压缩算法的失真低于某一视觉失真阈值时,对于编码重建后的图象,人眼觉察不到失真,则该阈值被称为刚能鉴别的失真阈值。利用了 JND 的概念, Jayant 归纳出基于人眼感知特性进行图象编码的框架,如图 2 所示。

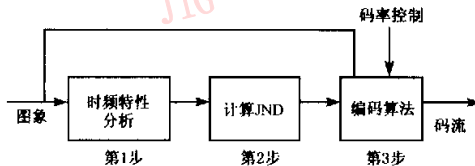


图 2 基于视觉特性编码的框架

输入图象在第 1 步中进行时频分析,抽取空间频率、背景亮度、纹理等图象特征。这些特征在第 2 步中用以计算视觉失真阈值。以失真不高于 JND 为准则在第 3 步中对图象进行压缩编码。

文献[9]最早提出了利用视觉感知特性进行编码,文中证明了用 MSE 这样简单的失真评估方式不能真实地体现出一幅图象与另一幅图象在视觉上的差异。根据从心理视觉实验得知的一些人眼视觉系统特性, Mannos 和 Sakrison 提出了对比敏感性函数(contrast sensitivity function CSF)的近似表达式,如图 3 所示,CSF 刻画了视觉系统对不同的

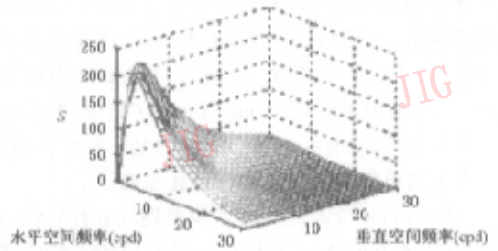


图 3 对比敏感性函数曲线

2D 空间频率有不同的敏感度。

这种 CSF 模式被结合到失真评估中,可以更好地预测图象的感知质量。在文献[9]基础上,研究者们作了更深入的研究,提出了多种基于视觉感知特性的编码方法。文献[10]扩展了文献[4]中的方法,提出了一种用于图象编码的加权余弦变换。为了使 CSF 适用于离散余弦变换 DCT,该文提出了一种简单的 CSF 模式,此模式用在人眼的最小聚焦视距下测出各频带的 JND 阈值来表示。这些阈值可结合到基于 DCT 编码器中,用于量化和比特分配^[4,11]。但其观测 JND 阈值的方法过于简单,使得 CSF 模式比较粗糙,因此 Saghri 通过计算显示器的校正、观测视距和图象分辨率更进一步地精细了文献[10]的方法^[12]。Eggerton 和 Srinath 则提出了一种基于 DCT 编码的感知加权的量化方式^[13]。Macq 导出:在变换编码中,感知加权因子与变换系数的量化噪音有关^[14]。他推导出的权重因子是一个关于显示和观测条件的函数,但与图象内容无关。

上述编码方法中仅用到了视觉的频率特性,而没有用到视觉对比和掩盖效应。频率特性是全局特性,仅与图象的大小和观测条件有关,而与图象的内容无关。为了更精确的计算人眼容忍失真的视觉阈值,还需要考虑视觉对比和掩盖效应。视觉对比和掩盖效应与图象内容有关。换句话说,精确的自适应的失真视觉阈值与人眼视觉的频率特性(MTF 曲线)、图象对比灵敏度 and 局部纹理的描述相关。要在给定的比特下达到最好的恢复图象效果,需要考虑上述所有的视觉特性。Chen 应用固定量化矩阵的观点来说明视觉系统的 MTF 特性,在编码中只对超出局部阈值的部分进行编码^[15]。他提出的编码器结构被认为是同期基于视觉感知编码的原型。后来, Chitprasert, McLaren, Ngan 等人在编码器中利用了频率、对比敏感性以及不同形式的纹理掩盖^[16~18]。

在这些编码器中并没有直接计算人眼可容忍的失真视觉阈值,而是结合频率、纹理、对比这些特性来对每个 DCT 块进行分类,通过分类结果来确定使用适当的量化规则. Safranek 则综合了频率、背景亮度和纹理的影响,直接计算可容忍的最小视觉阈值后,利用该阈值对子带图象进行量化编码^[19].

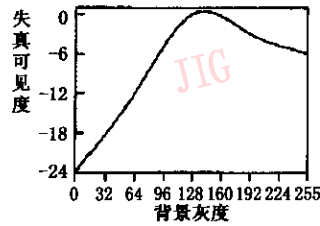
Safranek 等人的方法^[19]是早期具有代表性的基于 JND 阈值编码的方法之一. 他们将失真敏感性推导为一个关于频率、亮度、纹理的函数. 在计算 JND 阈值之前,先将图象通过镜像滤波器组后分解成 16 个频率子带 $s(s=0, \dots, 15)$, 如图 4(a)所示. JND 阈值的计算需要以下 3 个步骤:首先,要获取基本的频

0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

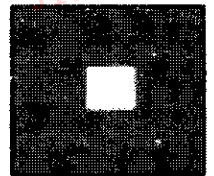
(a) 16 个频率子带编号

0.25	0.4	2.0	6.0
0.5	1.0	4.0	8.0
20	3.0	4.0	6.0
3.0	6.0	10.0	11.0

(b) 各频率子带的基本视觉失真阈



(c) 不同背景亮度下的失真敏感度调



(d) 背景图象中加入噪声块

图 4 Safranek 压缩编码方法图例

在编码时, Safranek 方法对编码图象的每个频率子带计算一个不同的的量化阈值. 由于对同一频率子带中所有的系数都使用固定的同一个量化步长,不能通过掩盖效应来适应图象的局部变化,因此往往会对人眼不敏感的区域分配过多的比特. 对此, Tran 和 Safranek 提出了自适应的感知编码方法^[20],应用图象分割的方法来达到局部自适应. 在自适应的感知编码方法中可以通过图象内容的变化来自适应地调节量化步长,但需要传输分割信息. Höntsch 提出不需要对图象进行分割而以 DCT 变换后的 8×8 大小系数块作为编码单位^[21],该方法可根据图象内容进行自适应感知编码,但存在如何协调块内对比掩盖效应与块间对比掩盖效应的问题.

小波分析(wavelet analysis)是傅立叶分析与调和分析发展史上的一个里程碑,被誉为“数学显微镜”. 作为一种多分辨分析方法^[22],由于小波变换具有很好的时频和空频局部特性,特别适合于按照人眼视觉系统特性设计的图象编码方案^[23~25]. 小波编码在压缩比和编码质量方面优于传统的 DCT 变换编码,已被应用到 MPEG-4 和 JPEG-2000 编码标准中. 小波变换将一幅图象分解成不同频带的子图. Watson 等人通过对背景亮度值为 128 的空白图象进行小波变换后,分别

率敏感视觉失真阈值,将所讨论的频率子带的亮度全部置为一定值 $I(I=128)$ 后加入功率可变的随机噪声,如图 4(d)所示,而后将加噪图象通过重构滤波器,让观察者在特定的距离观测重建图象,并确定在当前条件下刚能鉴别出的噪声值 B ,则 B 与 I 的差值就为该频率子带的基本视觉失真阈值,各频带的视觉失真阈值如图 4(b)所示;其次,变换背景亮度来测该频率子带在不同亮度下的失真敏感性,实验表明,人眼在中等灰度背景时,对噪声方块最敏感,亮度调节函数如图 4(c)所示;最后,纹理敏感性通过经验值来计算, Safranek 将纹理(Texture)定义为纹理能量及频率子带块内部的局部方差之和.

在小波系数的各频带子图上加入噪声,经过小波逆变换后,在空域上测出人眼对不同频率子图上的基本 JND 阈值^[26]. Höntsch 根据 Watson 测出的小波变换域中不同频带所能容忍的量化误差,提出了一种基于视觉感知特性的零树编码算法 PZEW^[27]. 该算法遵循人眼的视觉特性,以最小化视觉失真为准则对不同频带的小波系数加权后进行压缩编码,从而按视觉感知特性来分配比特,使得重建图象更符合人眼的主观感觉.

JPEG 和 JPEG-2000 标准中已将人眼视觉特性结合到压缩编码方法中, JPEG 中的量化表可允许对每个 8×8 的块进行频率和颜色加权,而 JPEG-2000 可支持更多的视觉调节工具^[28],如视觉渐进加权(Visual progressive weighting),邻域掩盖(Neighborhood masking)、点方式自掩盖(Point-wise self-masking)以及点方式扩展掩盖(Point-wise extended masking)等. JPEG-2000 中的视觉调节工具各有优缺点,但在一定的条件下,它们可以起到调节图象视觉效果的目的.

3 基于人眼感兴趣区域的编码

人类的视觉还具有一定的选择性. 人们在观察

和理解图象时能够不自觉地对其中的某些区域产生兴趣,把这些区域称为“视觉感兴趣区”。从生理解剖学可知,视网膜的作用是把外界的光学信号转变为生理反应,它直接感受外界光刺激,产生的生物电脉冲信号经过神经纤维传递到视神经中枢,最后在大脑中形成影像的感觉。视网膜的黄斑区集中了大量的锥状细胞,能够在强光下检测亮度和颜色信息,分辨率很高;而散布在其他区域的杆状细胞只能在弱光下检测亮度信息,没有色彩的感觉,分辨率较低,仅能呈现出视野中景物的一般轮廓。人在观察景物时,通过眼球的转动将目标“聚焦”在黄斑区上成像以看清物体。虽然人脑内部神经网络的计算是大规模并行的,但对于视觉输入信号的并行处理能力却是十分有限的。人无法同时识别同一复杂区域内的多个目标,只有进入黄斑区的景物信息才被充分接收,而其他的大量信息最终被中心滤除,即视觉受大脑的选择性注意机制控制。视觉的选择性与客体的特性也有很大的关系,人眼对空间频率接近于零的平滑区域和空间频率相似的纹理区域有很大的“钝性”,所以“不变”和“规则变化”的场景很容易在人的意识中被遗忘,人类视觉常常只对“突变”和“极不规则变化”的区域感兴趣。这种“与生俱来”的选择性使视觉只限定在有限的目标上,避免了过多精力的浪费,并在一定程度上起到了信息压缩的作用。

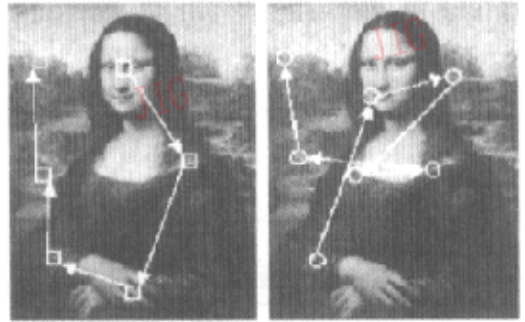
针对人眼视觉的选择性,人们提出了基于感兴趣区域的编码方案^[29~34]。对处理的图象中用户感兴趣的区域(ROI)进行压缩时采用低压缩比,而对感兴趣之外的区域采用高压缩比。在保证不丢失重要信息的同时,又能有效的压缩数据量,这就是基于感兴趣区域的编码方案所采取的压缩策略。

3.1 感兴趣区域(ROI)的确定

感兴趣区域的选择有两种方式^[35],一种是通过人根据其感兴趣的区域选定,称为人为检测的感兴趣区域 hROI (human detected regions-of-Interest);另一种是根据被处理图象的特征利用算法获得的感兴趣区域 aROI (algorithmically detected regions-of-Interest)。Stark 通过实验推出,人在识别一幅图象时,会移动视线寻找一些感兴趣的点,从而达到对该图象的识别^[36]。如图 5 所示,图中(a)是人眼的注意点; (b)是测试出的人眼聚焦点以及扫描次序; (c)是通过简单的权值算子推算出的注意点以及扫描次序。实验结果表明,人眼的注意点都是图象中一些发生“突变”的区域。



(a) 人眼的注意点



(b) 测出的人眼聚焦点及次序

(c) 算出的聚焦点及次序

图 5 Stark 测出的人眼注意点及扫描次序

目前,ROI 的计算方法基本上是以 Stark 的实验结果为基础,通过提取特征计算出图象中“突变”和“极不规则变化”的区域,并将这些区域确定为感兴趣的区域。Privitera 通过 10 种算子来计算在一幅图象中人眼感兴趣的点,并通过聚类的方法将近距离的兴趣点聚成感兴趣区域(如图 6 所示)^[35]。



图 6 聚类得到的 ROI

在量化编码时根据人眼将注意力聚集在感兴趣点时,在该点附近一定范围内,人眼的分辨力很高,而对远离该点的区域,人眼的分辨力较低,

Privitera 对 JPEG 编码方法进行了改进,提出了焦点 JPEG 压缩编码(Focused JPEG),在编码中,对 8×8 图象块进行量化时,根据像素块与 aROI 的距离进行量化. Privitera 将 10 种算子计算出的感兴趣区域(ROI)与人眼检测的感兴趣区域进行比较,虽然没有一个算子计算出的 aROI 与 hROI 完全相同,但大体基本相似.对于人眼扫描的次序则很难用算法测出.另外,每个算子的执行时间需要几分钟,且聚类方法也需要 10min 左右.

Itti 通过模拟注意力聚焦机制,提出了利用显著图来确定感兴趣的区域^[37],通过在每个位置上抽取的信息与剩余图象进行比较来计算具有突出性的部分.注意力聚焦(Focus-of-attention)模型模拟生物视觉系统能快速检测和定位静态视网膜图象上“感兴趣”部分的能力,以便减少目标识别的数据量.注意力聚焦机制主要分为 3 个阶段,如图 7 所示.

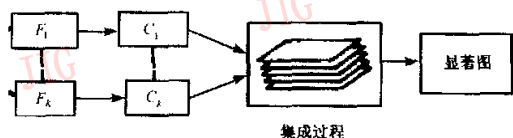


图 7 显著图的注意力聚焦机制

第 1 阶段,多特征图(Feature Maps, F-Maps) $F_k(k=1, \dots, K)$ 的获得,它从输入图象抽取,并反映了一些在视觉皮层计算出的图象特征.将输入图象经过红-绿-蓝-黄色彩对半的滤波器得到代表色彩的信息,其他一些特征图则代表无色的高频率信息,它们可通过一阶高斯滤波器得到.此文中抽取了 3 个特征图:亮度(I)特征图、颜色特征图(由红绿对比差值(RG)和蓝黄对比差值(BY)表示)、方向特征图(O);第 2 阶段,对关注图(Conspicuity Maps, C-maps) C_k 的抽取,对应于每个特征类型 k 的特征图抽取一个关注图.关注图代表在图象每个位置上有趣部分的 K 个从下往上的测度,它的值在 $[0, 1]$ 区间,这些测度可通过特征图与高斯滤波器的多尺度卷积得到,关注图 C_k 则通过计算在每个位置上的测度的平方得到.每个关注图突出那些在特定特征图中与其周围差异很大的区域;第 3 阶段,所有关注图 C_k 通过神经网络用竞争的方式合成成为一个显著图(Saliency Map, S-map).显著图的最大值表示了最引人注意的区域,把这些区域定为 ROI,可以用到压缩编码中. Ouerhani 将 Itti 的显著图确定

ROI 的方法用到 JPEG 压缩编码中,在低比特压缩时取得了较好的主观质量^[38]. Itti 的基于显著突出特征图的方法可以快速的检测出交通信息图象中不同形状(如圆、三角、矩形等)、颜色(如红、绿、白、黑等)、纹理(字母、箭头、条纹、圆等).但对于相互关联的目标(如 T 型连接或线形连接的目标)则很难检测到.另外,在提取图象基本特征以及用神经网络的方法计算突出特征时有很大的计算量.

在计算感兴趣的区域方面,近几年来有许多研究人员做了大量的工作,并取得了一定的进展.这些算法^[35-40]基本上都是先抽取图象特征,然后根据这些特征来确定能吸引人注意力的区域.但这些算法还存在许多问题,如算法的执行速度与有效性(即检测的 aROI 是否与 hROI 相近)等,因此,虽然目前 JPEG-2000 中已加入基于感兴趣的编码方案,但其中 ROI 是由用户指定.

3.2 JPEG-2000 中的基于感兴趣区域的编码

在 JPEG-2000^[4,5]中用户可以在被处理的图象中指定感兴趣的区域,对这些区域进行压缩时可以指定特定的压缩质量,或在恢复时指定特定的解压缩要求,而对感兴趣之外的区域采用高压缩比.在保证不丢失重要信息的同时,又能有效的压缩数据量. JPEG-2000 中用于压缩带有 ROI 图象的方法称为 Maxshift 方式,Maxshift 方式的基本原理是按比例缩小与 ROI 无关的小波系数,因此与 ROI 相关的系数在进行位平面编码时排在背景系数的前面,这样在解码端,在一定的解码率下 ROI 系数也能得到充分的解码. Maxshift 方式支持任意形状的 ROI 编码,并且具有较小的复杂度;无需向解码端发送 ROI 的形状信息,在解码端也无需对 ROI 进行特殊的考虑.图 8 中显示了基于指定 ROI 的编码重建效果图.



图 8 基于感兴趣区域编码效果图

4 基于轮廓边缘和纹理的压缩方法

基于视觉特性编码的另一种途径是根据图象分割的思想,对不同的实际分割块使用不同的压缩编码方法。早期这种方法被认为是子带编码的雏形。Schreiber 等人是最早尝试将图象分割后进行编码的^[41]。他们将图象分成高通信号和低通信号,然后对每部分独立编码。Yan 和 Sakrison 使用了相似的方法,将图象分成两个不同的部分,并分别进行压缩编码^[41]。该方法中使用了“轮廓边缘”概念,对于观察者来说,“轮廓边缘”可表示最重要的信息;假如对轮廓边缘进行了高质量的编码,则即使对图象的其余部分用很高的压缩比进行编码,整体图象的失真也不会很严重^[41]。对比分割后编码以及不分割压缩编码的重建图象结果后,Yan 和 Sakrison 得出结论:分割后编码的重建图象主观质量明显地高于未分割压缩编码的重建图象的主观质量。Faralle 和 Jain 提出了另一种相似的方法称为“递归块编码(recursive block coding)”^[42]。他们同样将图象分成两部分;一部分称为“边界块”,实际上是图象的二次抽样版本;另一部分是原图象减去边缘响应后的剩余部分。边界块用较高的信噪比编码,而相应的残余块则用二维 DCT 变换来编码。Faralle 和 Jain 分析了递归块编码的率失真(rate-distortion),从理论和实际重建图象结果上都证明了递归块编码优于标准的 DCT 编码方法。其主要的改进在于边缘块的误差很小,从而有效地消除了抑制效应(blocking effect)。抑制效应主要是由边缘误差引起的,人的视觉系统对直线和轮廓边缘极为敏感,因此,即使边缘块误差很小,人眼依然可以很明显地觉察到。

到了 20 世纪 80 年代,人们试图在基于视觉特性的分割编码基础上发展一种有别于第一代编码技术的图象压缩编码技术。这一技术通过识别图象的一些特征,并用这些特征来达到压缩的目的,该编码技术被称为第二代图象编码技术^[43]。第二代图象编码技术将人眼视觉系统特性结合到编码方法中,以便当压缩比很高时,图象仍然能保持人眼可接受的图象质量。大部分的第二代图象编码方法都是有损压缩。第二代图象编码技术,也称为结构编码,是指在编码时首先将图象中的边界、轮廓、纹理等结构特征提取出来予以保存,解码时根据结构特征参数合成图象^[43~44]。如第二代图象编码技术中的基于分割

的编码方法^[45,46](segmentation based)和基于草图^[47](sketch-based)的编码方法都是首先从图象中抽取相应区域的形状信息和纹理信息,产生“草图”,而纹理则由原图象和草图的差来定义,并使用波形编码技术进行编码。Ran 等人对该方法进行了扩展,提出了三分量图象模型,既将一个图象分成强边缘、纹理和光滑分量 3 部分^[48~50],其中纹理和光滑分量的编码采用波形编码,而对强边缘采用不同的编码方法,如链码方法进行编码。这种方法较好地保存了对人眼十分重要的边缘轮廓信息,因此在压缩比很高时解码图象质量仍然很好。随着小波变换在图象处理中的广泛应用,Mallet 等人提出了基于多尺度边缘检测的方法^[51],用于寻找最重要的边缘特征,并用边缘特征对图象进行压缩编码^[52]。通常,第二代图象压缩编码技术在压缩比高于 50 倍时,通常图象恢复效果的主观质量要比其他方法更好。目前对该项技术的研究正成为图象编码技术领域的一个热点问题,MPEG-4 VTC 算法^[53]的编码机制中体现了第二代图象压缩编码技术的思想。随着人们对多媒体信息的应用要求由播放型转向基于内容的访问和操作型,需要将基于内容的检索与编码结合起来考虑,在压缩数据中应有描述图象内容的信息,从而使对图象内容的访问可以直接针对压缩数据进行。因此第二代压缩编码技术在对图象进行分割时,不仅要利用方块内各像素间的统计特性,而且应当按照景物中物体形状进行分割,最终使图象按分割出来的对象进行编码,以便支持基于内容的交互性,第二代静态图象压缩编码技术也可称为基于内容的压缩编码技术。

第二代图象编码技术所面临的一个困难在于它通常所依赖的图象分割技术的不完善。图象分割技术目前虽然对一些特定的图象分割能得到很好的效果,但对复杂环境下或任意图象的自动分割还没有一个比较令人满意的方法^[44],因此这在一定程度上限制了第二代图象编码技术的发展。

5 结语与展望

如何使压缩图象的质量符合人眼的视觉特性,怎样将人眼的视觉特性结合到编码技术中是图象编码中的热点问题,它不仅与编码技术的发展有关,同时也与对人眼视觉系统的结构的认识紧密相关。基于人眼视觉特性的编码要求是在提高压缩比的同

时,又不损害压缩图象的主观质量,经过几十年的发展,关于基于人眼视觉特性的编码已提出许多成熟的技术,并应用到实际的编码中,如已应用于高清晰度电视的编码和传送、JPEG 压缩编码和 JPEG-2000 标准以及 MPEG-4 中,但到目前为止基于人眼视觉特性的编码技术还存在许多问题,有待于进一步解决,这些问题包括:

(1) 如何评价基于视觉特性压缩编码的重建图象的质量 一个压缩编码系统中的解压缩图象质量最终决定了该压缩技术的优劣,可以说,图象质量评价的实际意义就是对该系统及其技术质量的评价,传统的图象质量评价方法可分为客观和主观两类,客观的评价方法如均方误差 MSE,峰值信噪比 PSNR,虽然具有简单、易于实现等优点,但由于其根本不考虑人类的视觉和心理感受,不能反映人的视觉特性,且与人的主观评价脱节;主观的评价方法根据规定的评价标准和尺度,通过人眼直接观察图象打分评级,经统计平均得到的质量等级与实际情况比较吻合,但不易定量准确测量,带有明显主观性,受观察者本身的知识背景、情绪以及疲劳程度等因素的影响,且实现过程繁琐、耗时耗力耗资,因此,寻找基于人的视觉感知的图象质量评价方法显得极为必要,基于人眼视觉特性的图象质量评价方法的研究是基于视觉特性压缩编码技术进一步发展的基石。

(2) 对感兴趣区域的确定 虽然在 JPEG-2000 中引入了基于感兴趣区域的压缩编码方法,但需要通过用户来人为确定图象中感兴趣区域,这样就限制了该技术的应用,如何实时地自动确定图象中人眼感兴趣区域,对于基于感兴趣区域压缩编码技术的进一步应用有很重要的意义;另外,如果测试图象中包含多个感兴趣区,如何确定这些感兴趣区域的优先级? 这些问题依然有待于进一步的解决。

(3) 对象的识别和分割算法 MPEG-4 中已经体现出基于轮廓和纹理第二代压缩编码算法的思想,虽然 MPEG-4 标准的主要内容、编码工具以及码流格式已经基本上确定,并已有不少商品化的 MPEG-4 编解码器,但这些 MPEG-4 软硬件编解码器都不具有自动分割任意视频运动对象的功能,不能提供基于对象的操作功能,它们主要还是采用基于矩形块的编码技术,同时结合各种分级编码和传输技术以便在图象质量和码率之间做出平衡,没有实现真正意义上的基于内容的压缩编码,第二代图

象编码技术所面临的一个困难是它通常所依赖的图象分割技术目前虽然对一些特定的图象分割能得到很好的效果,但对复杂环境下或任意图象的自动分割还有待于进一步的研究。

基于视觉特性的编码方法依赖于对视觉系统的进一步研究,在对人眼视觉特性,特别是视觉心理特性进行深入研究的基础上结合先进的图象分析和图象理解知识,创建出更加符合科学的视觉模型,从而使基于视觉特性的编码技术的编码效果真正与人眼的视觉要求一致,随着多媒体技术的发展与网络的普及,基于视觉特性的编码技术将具有更广阔的应用前景和发展空间。

参考文献

- 1 Shreiber W F, Knapp C F, Kay N D. Synthetic highs: An experimental TV bandwidth reduction system [J]. Society of Motion Picture and Television Engineers Journal, 1959, 68(8): 525~537.
- 2 Netravali A, Petajan E, Knauer S. A high quality digital HDTV codec [J]. IEEE Transactions on Consumer Electronics, 1991, 37(3): 320~330.
- 3 Wallace G K. The JPEG still picture compression standard [J]. Communication of the ACM, 1991, 34(4): 34~43.
- 4 Askel J, Carlander M L, Christopoulos C. Region of interest coding in JPEG 2000 [J]. Signal Processing Image Communication, 2002, 17(1): 105~111.
- 5 Skodras A N, Christopoulos C A, Ebrahimi T. JPEG2000: The upcoming still image compression standard [J]. Pattern Recognition Letters, 2001, 22(12): 1337~1345.
- 6 Touradj E, Caspar H. MPEG 4 natural video coding: An overview [J]. Signal Processing, Image Communication, 2000, 15(1): 365~385.
- 7 Forchheimer R, Kronander T. Image coding: from waveforms to animation [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1989, 37(12): 2008~2023.
- 8 Jayant N. Signal compression based on models of human perception [J]. Proceedings of the IEEE, 1993, 81(10): 1385~1422.
- 9 Mamos J L, Sakrison J D. The effects of a visual fidelity criterion on the encoding of images [J]. IEEE Transactions on Information Theory, 1974, 20(4): 525~536.
- 10 Nill N B. A visual model weighted cosine transform for image compression and quality assessment [J]. IEEE Transactions on Communications, 1985, 33(6): 551.
- 11 Perkins M G, Lookabaugh T. A psychophysically justified bit allocation algorithm for subband image coding system [A]. In: Proceedings of IEEE ICASSP-89 [C], Glasgow, UK, 1989, 3: 1815~1818.

- 12 Saghri J A, Cheatham P S, Habibi A. Image quality measure based on a human visual system model[J]. *Optical Engineering*, 1989, 28(7): 813~818.
- 13 Eggerton J D, Srinath M D. A visually weighted quantization scheme for image bandwidth compression at low data rate[J]. *IEEE Transactions on Communications*, 1986, 34(8): 840~847.
- 14 Macq B. Weighted optimum bit allocations to orthogonal transforms for picture coding [J]. *IEEE Selected Areas in Communications*, 1992, 10(6): 875~883.
- 15 Chen W H, Pratt W K. Scene adaptive coder [J]. *IEEE Transactions on Communications*, 1984, 32(5): 225~232.
- 16 Chitrasert B, Rao K R. Human visual weighted progressive image transmission[J]. *IEEE Transactions on Communications*, 1990, 38(7): 1040~1044.
- 17 McLaren D L, Nguyen D T. Removal of subjective redundancy from DCT-coded images [J]. *IEE Proceedings - Communications, Speech and Vision*, 1991, 138(5): 345~350.
- 18 Ngan K N, Leong K S, Singh H. Adaptive cosine transform coding of image in perceptual domain[J]. *IEEE Transactions on Acoustics Speech and Signal Processing*, 1989, 37(11): 1743~1750.
- 19 Safranek R J, Johnston J D. A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression [A]. In: *Proceedings of IEEE ICASSP-89[C]*, Glasgow, UK, 1989, 3: 1945~1948.
- 20 Tran T D, Safranek R. A locally adaptive perceptual masking threshold model for image coding[A]. In: *Proceedings of IEEE ICASSP-96[C]*, Atlanta, GA, USA, 1996, 4: 1882~1885.
- 21 Höntsch I, Karam L J. Locally adaptive perceptual image coding [J]. *IEEE Transactions on Image Processing*, 2000, 9(9): 1472~1483.
- 22 Mallat S. Multifrequency channel decompositions of image and wavelet models [J]. *IEEE Transactions on Acoustics, Speech Signal Processing*, 1989, 37(12): 2091~2110.
- 23 Shapiro J M. Embedded image coding using zerotrees of wavelet coefficients[J]. *IEEE Transactions on Signal Processing*, 1993, 41(12): 3445~3462.
- 24 Lewi S A, Sknowles G. Image compression Using the 2-D wavelet transform[J]. *IEEE Transactions on Image Processing*, 1992, 1(4): 244~250.
- 25 Said A, Pearlman W A. A new fast and efficient image codec based on set partitioning in Hierarchical Trees [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 1996, 6(3): 243~250.
- 26 Watson A B. Visibility of wavelet Quantization noise[J]. *IEEE Transactions on Image Processing*, 1997, 6(8): 1165~1175.
- 27 Höntsch I, Karam L J. A perceptually tuned embedded zerotree image coder[A]. In: *IEEE International Conference on Image Processing[C]*, Santa Barbara, CA, USA, 1997: 41~44.
- 28 Zeng W J, Scott D. Shaw in Lei. An overview of the visual optimization tools in JPEG 2000[J]. *Signal Processing: Image Communication*, 2002, 17(1): 85~104.
- 29 Azizi S, Cochran D, McDonald J N. A sampling approach to region selective image compression[A]. In: *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers_C*, Pacific Grove, CA, USA, 2000, 2: 1063~1067.
- 30 Zhou W, Bovik A C. Embedded foveation image coding [J]. *IEEE Transactions on Image Processing*, 2001, 10(10): 1397~1411.
- 31 Park K, Lee C S, Park H W. A multiresolutional coding method based on SPIHT[J]. *Signal Processing: Image Communication*, 2002, 17(7): 467~476.
- 32 Chen M J, Pan C W, Chi M C. Improved region of interest image coder and its application[A]. In: *International Conference on Consumer Electronics, Digest of Technical Papers_C*, Atlanta, Georgia, USA, 2002: 226~227.
- 33 Nister D, Christopoulos C. Lossless region of interest coding [J]. *Signal Processing*, 1999, 78(1): 1~17.
- 34 Ström J, Cosman P C. Medical image compression with lossless regions of interest[J]. *Signal Processing*, 1997, 59(6): 155~171.
- 35 Privitera C M, Stark L W. Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(9): 970~982.
- 36 Stark L W, Choi Y. Experimental Metaphysics: The Scanpath as an Epistemological Mechanism[A]. In: *Advance in Psychology: Visual Attention and Cognition[C]*, Amsterdam, Netherland; Elsevier Science, 1996: 3~69.
- 37 Itti L, Koch C, Niebur E. A Model of Saliency Based Visual Attention for Rapid Scene Analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(11): 1254~1259.
- 38 Ouerhani N, Bracamonte J, Hugli H. Adaptive color image compression based on visual attention [A]. In: *11th International Conference on Image Analysis and Processing [C]*, Palermo, Italy, 2001: 416~421.
- 39 Milanese R, Gil S, Pun T. Attentive Mechanisms for Dynamic and Static Scene Analysis[J]. *Optical Engineering*, 1995, 34(8): 2428~2434.
- 40 Tsotsos J K, Culhane S M. Modelling Visual Attention via Selective Tuning[J]. *Artificial Intelligence*, 1995, 78(10): 507~545.
- 41 Yan J K, Sakrison D J. Encoding of imaged based on a two component source model [J]. *IEEE Transactions on Communications*, 1977, 25(11): 1315~1322.
- 42 Farelle P M, Jain A K. Recursive block coding--A new approach to transform coding [J]. *IEEE Transactions on Communications*, 1986, 34(2): 161~179.
- 43 Kunt M, Ikononopoulos A, Kocher M. Second generation

- image coding[J]. Proceeding of IEEE. 1985,73(4):549~574.
- 44 Reid M M, Millar R J, Black N D. Second-Generation Image Coding: An Overview[J]. ACM Computing Surveys. 1997, 29(3):3~29.
- 45 Jang J, Rajala S. Segmentation based image coding using fractals and the human visual system [A]. In: Proceedings of IEEE ICASP-90[C], Albuquerque, NM, USA. 1990;525~536.
- 46 Jang J, Rajala S. Texture segmentation-based image coder incorporating properties of the human visual system[A]. In: Proceedings of IEEE ICASSP-91[C], Toronto, Ont., Canada, 1991,2753~2756.
- 47 Carlsson S. Sketch-based coding of grey level images[J]. Signal Processing. 1988. 15(1):57~83.
- 48 Ran X, Farvardin N. Adaptive DCT image coding on a three component image model[A]. In: Proc of IEEE ICASSP-92[C], San Francisco USA, 1992;201~204.
- 49 Ran X, Farvardin N. A perceptually motivated three-component image model-part I: description of the model [J]. IEEE Transactions on Image Processing, 1995, 4(4):401~415.
- 50 Ran X, Farvardin N. A perceptually motivated three-component image model-part II: applications to image compression [J]. IEEE Transactions on Image Processing, 1995, 4(4):430~447.
- 51 Mallat S, Zhong S. Characterization of signals from multiscale edges[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(7):710~732.
- 52 Mallat S, Zhong S. Compact coding from edges with wavelets [A]. In: Proceedings of IEEE ICASSP-91[C], Toronto, Ont., Canada, 1991;2745~2748.
- 53 Moccagatta I, Chen H. MPEG-4 visual texture coding: more than just compression [A]. In: Proceedings of the International Conference on Consumer Electronics[C], Chiba, Japan, 1999: 302~303.



沃 焱 1975年生,1999年毕业于兰州大学计算机系获理学硕士学位,现为华南理工大学计算机科学与工程学院博士研究生,主要研究方向为多媒体数据压缩与网络传输。



韩国强 1962年生,1988年毕业于中山大学计算机科学系获理学博士学位,现为华南理工大学计算机科学与工程学院教授,博士生导师,主要研究方向为多媒体数据压缩与网络传输,发表论文50余篇。



张 波 1974年生,1999年毕业于兰州大学数学系获理学硕士学位,现为广东省中山市公安局软件工程师,主要研究方向为图象处理,计算机网络安全。