文章编号:1001 - 9081(2005)03 - 0589 - 03

一种低码率彩色人脸图像编码方法

李　倩,朱玉文,刘万春
(北京理工大学 计算机科学与工程系,北京 100081)
(xiaoqings22@ sina. com)

摘　要:提出了一种适合低码率信道传输的嵌入式彩色人脸图像编码方法,不仅把握了人脸图像的特点,而且充分利用了彩色图像小波变换后的多种相关性。该方法将联合色彩分量矢量量化和零块、零树编码方法巧妙结合,把彩色图像的三个色彩分量单独扫描,单独编码的过程简化为联合扫描和联合编码的过程。实验结果表明该算法在低码率下,相比 JPEG,获得了在视觉效果和峰值信噪比方面更好的恢复图像质量。

关键词:彩色人脸图像编码;联合色彩分量矢量量化;零树;零块
中图分类号: TP391.41　**文献标识码**:A

Color facial image coding for low bit rate transmission

LI Qian, ZHU Yu-wen, LIU Wan-chun
(*Department of Computer Science and Technology, Beijing Institute of Technology, Beijing* 100081, *China*)

Abstract: An embedded color facial image coding, based on the characteristics of facial image and multi-correlations of color image with wavelet decomposition, was proposed for low bit rate transmission. This algorithm combined joint-component vector quantization respectively with zero block and zero tree coding, scanned three component of color image separately and simplified separate coding into joint scanning and joint encoding. Experiment results show that in low bit rate environment, the new image compression scheme performs better than JPEG in the aspects of perception and PSNR(Peak Signal Noise Ratio).

Key words: color facial image coding; joint-component vector quantization; zero tree; zero block

0　引言

随着当前视频会议、可视电话等低带宽通信的广泛应用,低码率下图像压缩技术的研究也迅速展开。人脸图像作为其中最常见的一种图像模式,引起了人们的广泛关注,出现了专门研究这类特定图像的压缩算法[1,2]。这些算法普遍采用基于感兴趣区的编码方法,将整幅人脸图像分为感兴趣区(人脸区)和非感兴趣区(背景区),在编码时通过对不同区域采用不同精度的编码来获得较好主观质量的恢复图像,重点就在于尽可能减少感兴趣区的失真。但这些算法在对感兴趣区的编码中都没有很好利用人脸的自身统计特性,因而算法并不高效。本文提出了一种高效的人脸图像压缩算法:在感兴趣区的编码中考虑到了人脸不同区域的统计特性,采用了分类多级矢量量化方法;在非感兴趣区的编码中利用矢量量化压缩比大的特点进行快速编码,取得了很好的压缩效果。

在具体算法的设计上,尽管小波变换后小波系数之间的线性相关性很小,但仍然有相当数量的非线性相关性残留在系数之间,如何有效地利用这些相关性是图像压缩的关键。依据所采用的量化策略,小波图像编码可分为标量量化和矢量量化两类方法。在标量量化编码方法中,采用了较为有效的零树[3,4]、零块[5]结构消除了一部分相关性,取得了较好的编码性能。缺点是若想进一步消除小波变换后残留的其他相关性很困难,而矢量量化可以弥补这个缺陷。由于矢量量化能有效利用矢量中各分量间的相关性,所以如果将零树、零块结构矢量化,就可以很容易地挖掘到更多的相关性,编码的效果也会进一步得到提升。

彩色人脸图像经过多级小波变换后,由于最低频子带集中了图像的绝大部分能量,所以即使它的损失较小,也可能会较大地影响复原图像的质量。因此将最低频子带与其他高频子带分离,分别进行编码。高频子带是压缩算法的重点。为区分感兴趣区和非感兴趣区,这里引用了经典的基于神经网络的人脸检测算法[6]并结合肤色检测来定位人脸,得到一个不规则的人脸掩膜。图 1 为人脸掩膜经 3 级小波变换后的高频结构图,每个子带被分成了两部分(F_{mo},B_{mo}),F 代表人脸,B 代表背景,$m = 1,2,\cdots,M$,代表表示尺度属性,$o = 1,2,3$,表示方向属性(水平、垂直、对角线),同一尺度下 $F_i = F_{i1} + F_{i2} + F_{i3}$,$B_i = B_{i1} + B_{i2} + B_{i3}$。

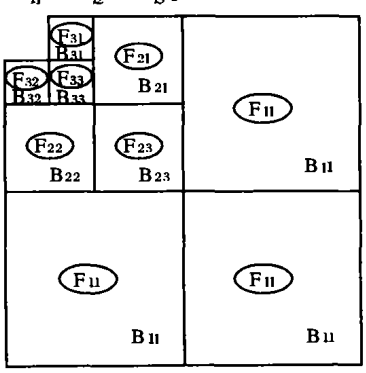


图 1　人脸掩膜经 3 级小波变换后的高频示意图

收稿日期:2004 - 08 - 24

作者简介:李倩(1979 -),女,湖南常德人,硕士研究生,主要研究方向:图像编码;　朱玉文(1950 -),女,四川成都人,副教授,主要研究方向:图像处理、人工智能;　刘万春(1949 -),男,河北滦县人,教授,主要研究方向:模式识别.

1 矢量组织

小波变换后,小波系数之间仍然有相当数量的非线性相关性,主要包括以下几类:1)同一方向不同尺度的子带之间具有偏序性和自相似性;2)不同方向同一尺度子带之间具有相关性;3)子带内的相邻系数之间也具有相关性。尽可能将具有相关性的小波系数组织在一个矢量中,通过矢量量化自动降低矢量各分量之间的相关性是本文进行矢量组织的出发点。要说明的是由于零树和零块结构能够有效地消除小波变换后的第一类相关性,矢量组织主要是针对消除小波变换后的后两类相关性以及彩色图像所特有的色彩分量间很强的频谱相关性。

这里的色彩空间选择了 RGB,具体表示如下:对于单个色彩平面 $X, X \in \{R,G,B\}$,针对小波变换后的第三类相关性,将各子带按照块大小 s^2 进行分块,每个块左上角系数的坐标记为此块的坐标;考虑到小波变换后的第二类相关性,然后将同一尺度 $m(1 \leqslant m \leqslant M)$ 下的 3 个代表不同方向的子带中对应 LL_m 中相同位置 (i,j) 的块按照 HL_m, LH_m, HH_m 顺序组织成一个维数为 $s^2 \times 3$ 的单平面矢量,记为 $V_X(i,j,s,m)$,例如图 2 中的黑色区域代表一个矢量 $V_X(0,0,2,1)$;最后为了能够消除色彩分量间频谱相关性,将相同位置的 R,G,B 单平面矢量组合起来形成一个 $s^2 \times 3 \times 3$ 维的联合色彩分量矢量 $\boldsymbol{V}(i,j,s,m)$。

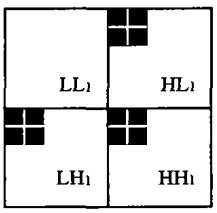


图 2 黑色区域代表矢量 $V_X(0,0,2,1)$

2 扫描图像

量化编码中的一个重要过程就是扫描。在矢量量化编码方法中,扫描是指按照某种设定的遍历次序以矢量为单位,判断此矢量对于当前阈值是否显著,并标识其状态;然后根据不同的状态,对系数采用不同的量化策略。对于每一个矢量,只要其中有一个分量的绝对值大于等于阈值,则认为是显著的,反之则是非显著的。当矢量维数较大时,判断一个矢量的显著性会需要更多的时间。在对联合色彩分量的矢量扫描时,当找到和 RGB 三分量密切相关的组合特征量作为扫描单位时,这一组合特征量是显著的,那么很大程度上相同空间位置的联合色彩分量矢量也是显著的,则可以进一步减少扫描时间。这里采用了 R,G,B 的线性组合得到的亮度 $Y(Y = 0.30R + 0.586G + 0.113B)$ 作为扫描中的组合特征量。而且,根据人眼视觉特性,随着尺度的递减,人眼敏感度也明显降低。为突出低频信息的显著性,在进行扫描前对亮度图像 Y 中所有高频数据还进行了阶梯量化,得到最终的扫描图像,记为 Y'。

3 背景区编码

图 1 中背景区 $B = B_1 + B_2 + B_3$ 不是人们所关心的,可以充分利用矢量量化压缩比大的特点,设计一种简单快速的方法进行压缩。在这部分将简单的零树结构[3] 和矢量量化相结合,也就是说如果一个父矢量是非显著的,则对应的相同空间位置相同方向上的所有子孙矢量也都在很大程度上是非显著的,这时就形成零值矢量树,此时父矢量就是零值矢量树根。整个这棵零值矢量树不需要进行矢量量化,只需 1 位表示,大大提高了编码效率。矢量组织上选择块大小 $s = 2$,为了和人脸区的矢量相区别,背景区的单个色彩平面矢量用 $\boldsymbol{BV}_X$ 表示,背景区的联合色彩分量的矢量用 $\boldsymbol{BV}$ 来表示,$\boldsymbol{BV} = (\boldsymbol{BV}_R, \boldsymbol{BV}_G, \boldsymbol{BV}_B)$,其中 $X \in \{R,G,B,Y'\}$。为了实现快速编码,借助非迭代的判决规则预测零值矢量树,然后采用矢量量化编码其中的显著矢量。算法描述如下:

1) 初始化。将 Y' 中所有属于 B_3 的系数组织成矢量,依次加入队列中;

2) 非迭代扫描。

```
队列不为空时循环
{
  取队头矢量进行显著性测试(例如 BV_Y'(i,j,2,m));
  传输 1 位表示其显著性(1 位显著的,0 为非显著的);
  如果此矢量是显著的,
    提取相应位置的联合色彩分量的矢量,(例如与 BV_Y'(i,j,2,m) 相对应的就是 BV(i,j,2,m)) 进行矢量量化,并传输其索引;
    如果扫描未到最小尺度,则将此矢量的四个孩子矢量加入队尾;(例如 BV_Y'(i,j,2,m) 的四个孩子矢量为 BV_Y'(i,j,2,m-1), BV_Y'(i,j+2,2,m-1), BV_Y'(i+2,j,2,m-1), BV_Y'(i+2,j+2,2,m-1));
}
```

4 人脸区编码

图 1 中人脸区 $F = F_1 + F_2 + F_3$ 是图像主要的感兴趣区,在这部分利用矢量量化的码本对于特定图像的优势,尽可能保证失真较小。但采用矢量量化存在着一个问题:即随着矢量维数的增加,压缩比加大,同时失真也较大。这里采用了 $s = 1$ 的较小维度的矢量结构,人脸区的单个色彩平面矢量用 $\boldsymbol{FV}_X$ 表示,人脸区的联合色彩分量的矢量用 $\boldsymbol{FV}$ 来表示。采用这种矢量结构不能很好地挖掘同一子带内小波系数的相关性,而零块结构正好可以弥补。

集合分割嵌入块编码器(SPECK)[5] 是最有代表性的基于零块的编码算法,它利用基于块的集合分割来消除同一零块中的小波系数的相关性,达到数据压缩的目的。SPECK 不仅考虑小波变换后的第一类相关性,还充分利用小波变换后的第三类相关性。

本文在人脸区的编码算法采用了类似于 SPECK 算法的零块思想,通过集合分割技术实现数据压缩。与 SPECK 算法不同的是,集合不再以系数为单位,而是以矢量为单位。也就是说如果此集合经过显著性测试被认为是显著的,表明此集合中至少有一个矢量是显著的,否则就是非显著的。由于只处理属于人脸区的高频系数,所以初始化时集合 I 设为 $F = F_1 + F_2 + F_3$,如图 3(a) 所示(为了便于理解,图中所画为整个的高频系数),然后对集合 I 进行显著性测试,如果整个集合 I 被认为是不显著的,则把这个集合 I 用 1bit 表示,否则按图 3(b) 所示的分割方案把集合 I 分割成两个子集合:集合 S 和集合 I。如果集合 I 还是显著的,再进行一次分裂,得到新集合 S 和新集合 I。如此重复直到集合 I 变为空,如 3(c) 所示。对于每个集合 S,也要进行显著性测试。如果集合 S 是显著的并且只包括一个矢量,则提取其相同位置的联合色彩分量的矢量进行矢量量化。对于包含不止一个矢量的集合 S,若是不显著的,按包含不显著矢量的数目,由少到多安排在不重要集合列表 LIS 中;否则按图 4 所示的分割方案把集合 S 分割成四个子

集合(S_1,S_2,S_3,S_4),每个子集 S_i 又包括三部分(S_{i1},S_{i2},S_{i3})。对于每个子集合也都要进行显著性测试,如果是显著的,继续分裂,如此重复直到找出集合中所有的显著矢量,并直接进行矢量量化。

相比 SPECK 算法,本算法去掉了系数幅度精细化扫描的过程,而是相对应每一阈值,若找到显著矢量 FV_Y,直接提取其相同位置的联合色彩分量的矢量 FV,采用分类多级矢量量化来减少失真。多级矢量量化由多个连续的矢量量化阶段构成,每个阶段具有一个很小的码本,失真比较小。不仅如此,注意到人脸这一感兴趣区中的某些特征(如眼睛、鼻子、嘴巴)是视觉的最敏感处,而且分别包括眼睛、鼻子、嘴巴的上中下三个不同区域存在着各自的统计特性。当矢量量化以某一类(如眼睛)为训练对象时,其训练得到的码本可以充分融入该对象的统计特性。为尽可能减少失真,将人脸区分为三类,分别训练自己的码本。因此对于每个显著矢量 FV,都必须先判断所属类别,然后进行相应的多级矢量量化。

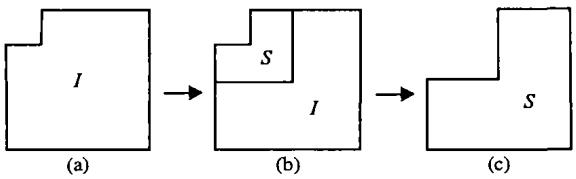


图 3　集合 I 的分裂方式

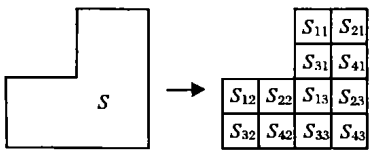


图 4　集合 S 的分裂方式

5　实验结果

这里所采用的人脸图像库共有包括不同光照和姿势下的 150 张真彩色图片,第 1 ~ 100 张被用来作为训练集以生成码本,剩下的图片(第101 ~ 150张)被用来测试。对于背景区,码本大小为 256;对于人脸区,码本大小为{16,8,8},均采用模糊竞争学习方法训练出以上码本。变换部分则采用了 9/7 双正交提升小波结构。

对图 5(a)(320 × 240)彩色图像(不属于训练集)进行了测试,在压缩比 CR = 75 的情况下,恢复图像的效果如图 5(b)所示,图 5(c)为 JPEG 算法的恢复图像,其中图 5(d)和图 5(e)分别为图 5(b),图 5(c)的人脸区部分的放大图。从主观上看,图 5(b)虽然背景比较模糊,但人脸表情等细节清晰可见,没有出现明显失真,而 JPEG 的恢复图像图 5(c)在人脸区已出现了严重的方块效应。从客观上看,表 1 比较了两者在不同压缩比 CR 下的峰值信噪比 PSNR。可以看到在同等压缩比的情况下,本算法重建图像人脸区的 PSNR 都远远高于 JPEG 重建图像人脸区的 PSNR。在中等压缩比时,本算法重建图像的整体 PSNR 偏低于 JPEG,这主要是针对低码率信道,对于背景区失真较大而造成的。但是当压缩比继续增大时,可以看到重建图像的整体 PSNR 值变化很平缓,而 JPEG 算法压缩恢复图像的整体 PSNR 值急剧下降。

可见,在高压缩比的情况下,本算法的压缩效果无论是从 PSNR 方面,还是从视觉效果方面来看,都优于 JPEG,特别适合于低码率下的人脸图像的压缩。

表 1　两种算法不同压缩比下 *PSNR* 值比较

对象	压缩比(CR)	PSNR	
		本算法	JPEG
整体	57	32.35	32.49
	75	30.29	30.36
	113	29.08	22.71
人脸	57	34.29	29.50
	75	33.15	27.47
	113	31.76	20.98
背景	57	31.99	33.19
	75	28.90	31.02
	113	28.89	23.04


(a) 原图像　(b) 本算法重构图像　(c) JPEG重构图像　(d) (b)的人脸区　(e) (c)的人脸区

图 5　实验结果图像

6　结语

本文充分利用矢量量化对于特定图像的优势以及视觉对于图像中不同感兴趣区的要求,对不同区域采用不同精度的编码策略。对于背景区,强调速度和压缩比;而对于人脸区,强调在保证了较高主观恢复质量的前提下尽可能提高压缩比。同时在传输上实现了渐进传输。实验结果证明了该方法的有效性,可广泛适用于可视电话、视频电视会议、人机交互等等低码率下主要传输人脸图像的编码领域。

参考文献:

[1] SAKALLI M, YAN H. Human face image coding using deformable templates, wavelet decomposition and VQ[J]. IEEE tencon, 1997, 33(4):569 - 572.

[2] SEOL J, SOHN K. Selective coding of human faces using wavelets [J]. IEEE Transactions on Image Processing, 2000, 6(8): 1384 - 1386.

[3] SHAPIRO JM. Embedded image coding using zerotrees of wavelet coefficients[J]. IEEE Transaction on Signal Processing, 1993, 41(12):3445 - 3463.

[4] SAID A, PEARLMAN WA. A new, fast and efficient image codec based on set partitioning in hierarchical trees[J]. IEEE Transaction on Circuit and Systems for Video Technology, 1996, 6(1): 243 - 250.

[5] PEARLAMN WA, ISLAM A. Low-complexity image coding with a set-partitioning embedded block coder[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 23(2): 1305 - 1311.

[6] ROWLEY H, KANADE T. Neural network-based face detection [J]. IEEE Transactions on Image Processing, 1996, 18(6): 203 - 208.