面向头肩序列图像的质量可精细扩展视频编码方法

卓 力[1], 沈兰荪[1], 林健文[2], 张延华[1]
(1. 北京工业大学信号与信息处理研究室, 北京 100022; 2. 香港理工大学电子及资讯工程系, 香港)

摘 要: FGS 编码方法具有细粒度的可扩展能力, 能很好地适应网络带宽的动态变化, 被认为是一种适合于网络视频传输的编码方案. 但现有的 MPEG-4 FGS 编码标准效率低, 限制了其进一步的推广应用. 因此, 本文面向视频应用中常见的头肩序列图像, 实现了一种质量可精细扩展的视频编码方法. 该方法采用 H. 26L 对基本层进行编码, 采用基于 DCT 变换的 SPIHT 方法对原始图像与基本层重建图像之间的残差进行编码得到增强层的码流. 然后将复杂背景下的人脸检测与跟踪技术与选择性增强技术结合起来, 对人脸区域优先编码. 实验结果表明, 该方法不仅编码效率高于现有的 MPEG-4 FGS 标准, 码流具有可精细扩展的特性, 还可以选择性地提高人脸区域重建图像的主观感受水平.

关键词: 头肩序列图像; 可精细扩展; H. 26L 编码; 选择性增强; SPIHT 算法
中图分类号: TN915.01 文献标识码: A 文章编号: 0372-2112 (2004) 03-0441-05

Quality Fine Granular Scalability Video Coding Method for Head Shoulder Sequence Images

ZHUO Li[1], SHEN Lan-sun[1], Kin-man LAM[2], ZHANG Yan-hua[1]
(1. Signal & Information Processing Lab, Beijing Polytechnic University, Beijing 100022, China;
2. Department of Electronic & Information Engineering, Hong Kong Polytechnic University, Hong Kong, China)

Abstract: Fine granular scalability (FGS) coding method has fine-grained scalable capability and can be adapted to dynamic variation of network bandwidth. So it is considered as a good coding scheme suitable for video transmission over network. But current MPEG-4 FGS coding standard is not efficient, which will restrict its application. Therefore, a quality fine granular scalable coding scheme for head shoulder sequence images, which are commonly seen in video applications, is implemented in this paper. The base layer is encoded with H. 26L and the residual signal between the original image and the reconstructed image from base layer is encoded with a DCT-based SPIHT coding method to achieve the enhancement layer bit stream. Automatic human face detection and tracing algorithm in a complex background is combined with selective enhancement technique to encode human face region with high priority. Experimental results show that the overall coding efficiency gain of this method is higher than that of MPEG-4 FGS standard and the bit stream is fine granular scalable. The subjective perceptual quality of reconstructed human face region can be selectively improved.

Key words: head shoulder sequence images; fine granular scalability; H. 26L; selective enhancement; SPIHT algorithm

1 引言

近年来, 随着计算机网络和第三代(3G)无线通信网络的不断发展, 视频传输已经成为多媒体信息服务中的重要内容, 视频编码的目标已由过去面向存储转向面向网络传输[1,2]. 为了提高在无线、因特网等复杂网络环境中的传输特性, 人们提出了细粒度可扩展(FGS, Fine Granular Scalability)编码方法[3~7]. 该方法将视频信息分成基本层和增强层, 基本层提供的是视频信号的基本信息, 允许单独解码, 可以满足网络传输带宽的最低要求, 通过它能提供基本的图像质量. 而增强层提供的是视频信号细节信息, 码流能够覆盖一定网络带宽变化的动态范围, 可以提供从基本层解码质量到近无损之间的可细粒度增强的图像质量. 细粒度意味着码流可以在任意处进行截断, 从而能够很好地适应网络带宽的动态变化. FGS 可以提供质量、时间和空间等多种可扩展方式. FGS 编码的基本思想也为视频传输提供了很好的解决方案, 因此成为面向网络传输的视频编码方法中的研究热点[3~7].

从现有的各种 FGS 编码方案看, 基本层编码时往往采用一种 DCT 变换+运动补偿的混合编码方法, 如 MPEG-4、MPEG-2 或者 H. 263 等, 而增强层编码时则可以采用任一细粒度的编码方法. 在 MPEG-4 FGS 编码标准中, 基本层采用单层 MPEG-4 编码方法, 而增强层采用比特平面(bit plane)编码方

收稿日期: 2002-12-23; 修回日期: 2003-08-18
基金项目: 国家自然科学基金(No. 90104013, 90304001); 863 计划(No. 2001AA121061, 2002AA123041); 北京市自然科学基金(No. KZ200310005004)

法对基本层重建图像与原始图像之间的残差进行编码[6]. Hayder 等则用 EZW(Embedded Zerotree Wavelet) 方法对增强层进行编码[3]. 这两种方法的编码效率都偏低, 因此影响了其进一步的实际应用. He 等将 H. 26L 与 MPEG2PFGS 结合起来, 在增强层中引入了运动估计补偿机制, 提出了一种高效的 FGS 编码方法, 在低、中、高三种码率下都可以获得比 MPEG24 FGS 更高的编码效率, 但这种方法获得的编码效率是以很高的编码复杂度为代价的[7].

MPEG24 FGS 标准提供了选择性增强和频率加权两种自适应量化技术来提高重建图像的主观感受水平. 选择性增强可以看成是一种特殊的比特分配技术, 通过它可以使得图像序列中某一被选定的区域能优先编码并传输, 这样增强层的码流在被截断时就有更多的比特分配给该区域, 从而能够有选择地增强该区域的重建图像质量. 需要说明的是, 选择性增强技术不能提高编码器的率失真性能, 而只能有效改善某一区域图像的主观感受质量[5]. 采用选择性增强技术时首先需要确定视频序列中人眼视觉感兴趣的区域, 但是对于大多数的视频序列来说, 确定图像中的人眼感兴趣区域是一件很困难的事情. 人脸是图像中最常见的视觉对象, 而头肩序列图像(Head Shoulder Sequence Images) 也是一种常见的一种图像模式, 在网络游戏、可视电话、视频消息、电视新闻与电视会议等视频业务中有广泛的应用.

为此, 本文提出了一种面向头肩序列图像高效的质量可扩展编码方法, 该方法首先采用 H. 26L 对基本层进行编码, 采用基于 DCT 变换的 SPIHT 算法(Set Partitioning In Hierarchical Trees) 对增强层进行编码, 该方法不仅实现简单, 复杂度低, 码流还具有细粒度可扩展的特性. 然后将复杂背景下的人脸检测与跟踪技术与选择性增强结合起来, 对图像中的人脸区域优先编码.

2 编码框架

本文方法的编码框架如图 1 所示. 从图 1 中可以看出, 本文提出的 FGS 编码方法主要包括三个部分: H. 26L 编码方法、基于 DCT 变换的 SPIHT 算法以及复杂背景下的人脸检测与跟踪算法. 下面分别介绍这三部分的内容.

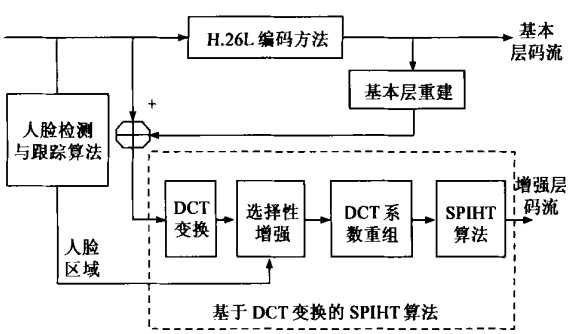

图 1 面向头肩序列图像的质量可扩展视频编码方法框图

2.1.1 H. 26L 编码方法

H. 26L 是由 ITU-T VCEG 发起并制定的视频编码标准, 与 MPEG24 等现有的国际标准相比, 其编码性能有了比较大的突破. 在 30-35dB 解码质量范围内, 最多可比 MPEG24 节约 50% 的码率. 目前 MPEG24 和 ITU 成立了 JVT 小组, 以 H. 26L 为基础, 联合制定 MPEG24 标准的 AVC 部分. 图 2 所示的是 H. 26L 的编码框图[8].

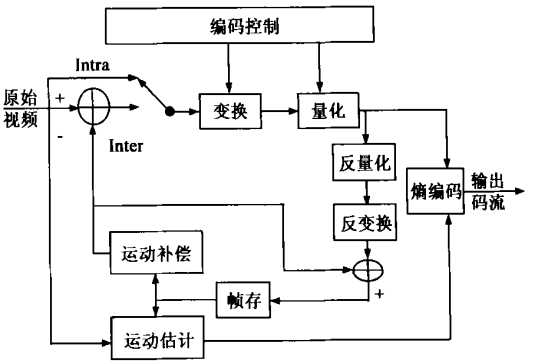

图 2 H. 26L 编码框图

从图 2 可以看出, H. 26L 仍然采用与 MPEG2 1/2/4 等现有的国际标准相同的基于变换和运动补偿预测的混合编码结构, 但不同的是它还采用了一些新的编码方法, 如 Intra 预测、7 种运动补偿块、4@4 整数变换、UVLC(Universal Variable Length Coding)、CABAC(Context-based Adaptive Binary Arithmetic Coding) 以及环滤波器(Loop Filter) 等. 通过采用这些技术, H. 26L以一定的编码复杂度为代价获得了编码效率的大幅提高. 有关 H. 26L 的详细内容可以参考文献[8].

2.1.2 基于 DCT 变换的 SPIHT 算法

SPIHT 算法是 A Said 和 Pearlman 在 EZW 算法的基础上提出的一种静止图像编码方法, 虽然 SPIHT 算法实现起来与 EZW 有较大差别, 但基本思想与 EZW 一致. 同 EZW 相比较, 它的主要成功之处还在于构造了如图 3 所示的空间方向树, 不仅可以充分利用不同尺度间小波系数的相关性, 还充分利用了同一尺度下相邻小波系数的相关性, 从而获得了比 EZW 更高的编码性能. 近年来, SPIHT 算法由于实现简单, 高效、运算复杂度低、码流具有天然的可精细扩展的特性等特点而被广泛应用于各种图像/视频编码方案中.

SPIHT 算法分别使用了三个列表来表示小波系数的编码状况, 即不重要系数表 LIP(the list of insignificant pixels)、重要系数表 LSP(the list of significant pixels) 和不重要集合表 LIS(the list of insignificant sets) .

SPIHT 算法的编码过程可以用如下的伪代码表示:

```
For b= bns, , , 0          从最重要的比特平面到最不重要的比特平面
Sorting Pass:
    Encode_Process_LIP;    搜索 LIP 列表中的重要系数;
    Encode_Process_LIS;    搜索 LIS 列表中的重要系数;
Refine Pass:
    Encode_Process_LSP;    细化 LSP 中的重要性系数;
```

Refine Pass 仅仅处理前一阈值下搜索出来的重要性系数, 对于当前阈值下搜索出来的重要性系数在 LSP 中暂不作细化处理, 而是等到下一阈值时再处理.

SPIHT 算法不仅可以对小波图像进行高效的压缩, 对于分块的 DCT 系数仍可以进行高效的压缩, 只要将变换后的 DCT 系数按照图 4 的方式进行重新组合, 将不同分块内相同位置处的 DCT 系数重新组合在一起就可以得到类似于小波图像的塔式数据结构. 图 5 所示的是 Lenna 图像 DCT 系数重组后得到的类似于小波图像的塔式数据结.

与基于小波变换的 SPIHT 算法相比, 基于 DCT 变换的 SPIHT 算法具有较低的运算复杂度, 在 1bpp 条件下得到的编码性能低 0. 7dB 左右, 但均高于 EZW 算法和 JPEG 标准[10, 11]. 考虑到小波变换的运算复杂度比较高, 同时避免在同一编码系统中出现两种不同的变换方式, 本文选择采用基于 DCT 变换的 SPIHT 算法对原始图像与基本层重建图像的残差进行编码得到增强层码流. 该方法不仅实现简单, 运算复杂度低, 码流还具有可精细扩展的特性[11].

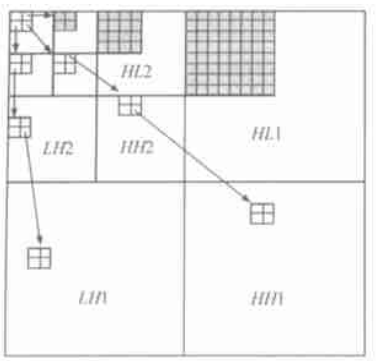

图 3　SPIHT 算法中的空间方向树

图 4　DCT 系数的重组顺序

图 5　Lenna 图像 DCT 系数重组后得到的塔式结构

213　人脸检测与跟踪算法

由于人类的视觉心理特性, 对于头肩序列来说, 一般情况下人眼对脸部区的感兴趣程度远高于肩部及其他背景区. 本文采用复杂背景下的人脸检测与跟踪算法来确定头肩序列图像中的人脸区域. 人脸检测用来从各种不同的场景中检测出人脸的存在并确定人脸区域, 人脸的位置是预先不知道的, 因而首先必须确定场景中是否存在人脸, 如果存在人脸, 再确定图像中人脸的位置.

不同人脸间具有共性, 即均由眉、眼、鼻和嘴等器官组成, 且具有固定的结构和良好的对称性. 本文采用的人脸检测算法正是充分利用人脸的这种独有的特点来进行人脸检测, 因此本文采用的人脸检测算法对光照、人脸姿势、表情、胡须、耳环以及除黑边眼镜、墨镜以外的眼镜等不敏感, 允许人脸有局部遮挡, 只要双眼和嘴等主要器官可见, 就可比较准确地检测出人脸, 因而可以获得很高的正确检测率[12,13]. 图 6 所示的是人脸检测的框图, 主要包括粗检、细检两个部分, 其中粗检包括颜色分割、脸部特征提取和候选脸检测等三个部分, 用来确定图像中所有可能的候选人脸; 而细检部分用于剔除错误检测的非人脸区域并从众多的候选脸中确定真正的人脸.

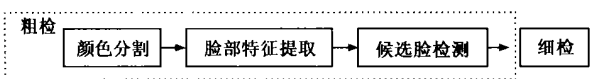

图 6　人脸检测框图

肤色是人脸的重要特征, 颜色分割就是用来分割出图像中可能的人脸区域. 肤色模型是在 YC_bC_r 空间中通过训练大量的肤色图像而得到的, 基于这一模型, 就可以粗略确定图像中的人脸区域. 接下来的脸部特征提取和候选脸检测就是利用人脸所特有的结构特征, 利用候选脸检测模板来确定图像中所有的候选脸区域. 这时得到的候选脸区域中可能包含有非人脸区域, 并且一个脸部区域可能出现多个候选脸. 细检部分的作用就是进一步剔除图像中的假脸, 合并同一区域的候选脸. 细检利用的是人脸区域灰度投影的特性: 整个人脸区域的水平投影至少有三个谷值点, 分别对应于人脸中的眼睛、鼻和嘴等位置, 而鼻尖和嘴唇区域则对应着两个峰值点. 眼睛部位的垂直投影有两个谷值点, 分别对应于两个眼睛. 利用这些特性就可以进一步正确确定人脸的真正位置.

人脸跟踪是根据已定位出的人脸, 在后续的运动图像序列中捕获该人脸的运动. 双眼、嘴等人脸器官特征明显, 而且在视频序列中, 它们的形状特征基本不变. 对于头肩序列图像来说, 人脸的运动一般较为缓慢. 为此本文采用的人脸跟踪方法首先利用前一帧图像中人脸检测确定的嘴部区域, 利用数学形态学上的运算快速分割出本帧图像中的嘴目标, 然后再确定整个的人脸区域[12,13]. 该方法不受背景颜色、光照等因素的影响, 具有很强的鲁棒性. 图 7 所示的分别是对 QCIF 格式 Foreman 序列第 3、9、13、17、23、27、31、40 帧的跟踪结果.

214　选择性增强技术

选择性增强技术用来实现人脸区域的优先编码, 目的是在低码率的情况下仍可以保证重建图像中人脸部分的质量能得到增强[5]. 在 MPEG24 FGS 编码方法中, 增强层编码采用的是比特平面编码的方法. 比特平面编码从最重要的非零比特平面开始, 直到最不重要的比特平面结束, 依次对比特平面进行编码. 因此每个系数的重要部分被优先编码, 按照重要性程度依次放置码流的其他部分, 这样在对码流进行截断时仍可以保证解码端能获得视频信息中最重要的部分, 提供了细粒度的可扩展能力.

对于 SPIHT 算法来说, 如果初始阈值选择的是 2 的幂次(如 512, 256 等), 则 SPIHT 算法本质上就是一种比特平面编码技术. 但由于 SPIHT 编码方法对每个系数最高平面以上的比特不予编码, 这样就编码效率来说, SPIHT 编码方法要高于一般的比特平面编码方法. 对于比特平面编码方法来说, 为了实现图像中的某一区域被优先编码并传输, 一个简单的方法就

是将该区域中的所有系数都上移若干个平面, 这相当于给该区域中每一个系数都乘以一个 2 的幂次方, 即

$$C'(i,j) = C(i,j) \# 2^{n(j)} \quad (1)$$

其中 $C(i,j)$ 表示的是第 j 个块中第 i 个系数的原始值, 通过式(1), 将第 j 块中的所有系数均上移 $n(j)$ 个平面.

如图 1 所示, 在编码端, 先对原始图像与基本层重建图像之间的残差图像进行 8* 8 的分块 DCT 变换, 然后对人脸区域的系数进行比特平面上移操作, 对得到的系数进行重组后再利用 SPIHT 编码算法进行编码. 在解码端, 先用 SPIHT 解码算法进行解码, 然后将 DCT 系数重新放置成原来的顺序, 对人脸区域的系数进行比特平面的下移操作, 再进行 DCT 反变换, 得到重建的残差图像.

需要指出的是, 选择性增强技术不能保证感兴趣区内的所有系数均被优先编码. 由于每个块的 $n(j)$ 数以及人脸区域的位置参数需要传送给解码端, 因此采用选择性增强技术还会在一定程度上造成编码效率的下降.

图 7 Foreman 序列的检测与跟踪结果(第 3、9、13、17、23、27、31、40 帧)

3 实验结果

为了验证本文提出方法的编码效率, 我们采用标准测试序列进行了实验. 测试序列包括 CIF 和 QCIF 两种格式, 帧率为 10fps, 编码长度为 10 秒. 图 8 所示的是采用本文方法和 MPEG24 FGS 方法分别对各测试序列编码后得到的比较结果. QCIF 格式序列的基本层码率为 32Kbps, CIF 格式序列的基本层码率为 128Kbps. 图 9 所示的是采用选择性增强技术之后得到的 Salesman 序列的重建图像, 其中基本层码率为 6kbps, 增强层码率为 12kbps. 图 10 所示的是采用选择性增强技术之后得到的 Foreman 序列的重建图像, 其中基本层码率为 50kbps, 增强层码率为 50kbps.

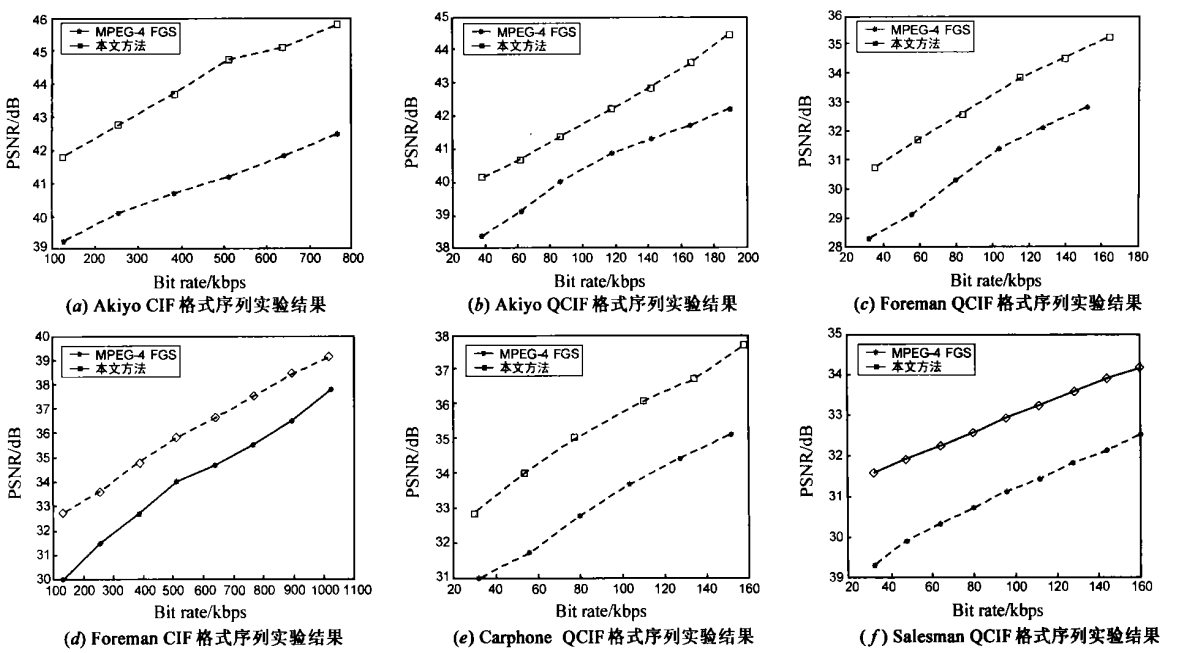

图 8 本文方法与 MPEG24 FGS 编码结果比较

图 9　QCIF 格式 Salesman 重建图像

图 10　CIF 格式 Foreman 重建图像

从图 8~ 10 可以看出, 本文方法的编码效率要比 MPEG2-4 FGS 平均好 2~ 3dB 左右, 采用选择性增强技术还可以使得人脸区的重建图像质量要好于背景区的重建图像质量.

在算法复杂度方面, 由于本文采用 H. 26L 作为基本层的编码方法, 因此整个算法的编码复杂度要远高于 MPEG2-4 FGS 编码方法, 而解码复杂度则基本相当. 可以说, 本文方法的编码效率是以一定程度的编码复杂度为代价的.

4　结论

FGS 编码方法具有细粒度的可扩展方式, 能很好地适应网络带宽的动态变化, 被认为是一种适合于网络视频传输的编码方案. 现有的 MPEG2-4 FGS 国际标准由于编码效率比较低, 因此影响了其进一步的推广应用, 如何提高细粒度可扩展编码的效率有待于深入的研究.

头肩序列图像是一种常见的图像模式, 广泛应用于各种视频应用中. 为此本文面向头肩序列图像, 提出了一种质量可精细扩展视频编码方法. 实验结果说明, 一方面该方法的编码效率要高于现有的国际标准 MPEG2-4 FGS, 可以提供细粒度的质量可扩展能力, 适应于在时变的无线、因特网等环境中传输. 另一方面结合复杂背景下的人脸检测与跟踪技术, 可以将有限的码率优先分配给图像中的人脸区域, 从而提高重建图像的主观感受水平.

如前所述, 由于基本层采用的是 H1 26L 的编码方法, 其算法的编码复杂度要高于 MPEG2-4 单层编码方法, 因此本文方法的编码效率是以一定程度的复杂度为代价的. 如何进一步降低算法的复杂度, 提高算法的实用性将是我们下一步的研究方向.

参考文献:

[1] 沈兰荪, 卓力, 等. 视频编码与低速率传输 [M]. 北京: 电子工业出版社, 2001.

[2] Da2peng WU, Yi2wei Thomas Hou, Ya2Qin ZHANG. Scalable video coding ad transport over brand2band wireless networks [J]. Proceed2ings of the IEEE, 2001, 89(1): 6- 20.

[3] Hayder Radha, et al. Scalable internet video using MPEG2-4 [J]. Signal Processing: Image Communication, 1999, 15: 95- 126.

[4] M van der Schaar, Radha H. Adaptive motion2compensation fine2granu2lar2scalability (AMC2FGS) for wireless video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(6): 360- 371.

[5] M van der Schaar, Y 2T Lin. Content based selective enhancement for streaming video [A]. Proc of 2001 Inter Conf on Image Processing [C]. Thessaloniki, Greece: IP, 2001, (2): 977- 980.

[6] We2ping LI. Overview of fine granularity scalability in MPEG2-4 video standard [J]. IEEE Trans Circuits and Systems for Video Technology, 2001, 11(3): 301- 317.

[7] Yu2wen HE, Feng WU, Shi2peng LI, Yu2zhuo ZHONG, Shi2qiang YANG. H. 26L2based fine granularity scalable video coding [A]. Proc of 2002 Inter Sym on Circuit and Systems [C]. Phoenix, USA: SCS, 2002(4): 548- 551.

[8] Kristofer Dovstam. Video Coding in H. 26L [OL]. http://www. cite2seer. nj. nec. com, Jan 2001.

[9] A Said, W A Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees [J]. IEEE Transaction on Circuit and System for Video Technology, 1996, 6(3): 243- 250.

[10] Zi2xiang XIONG, Kannan Ramchandran, Michael T Orchard, Ya2Qin Zhang. A comparative Study of DCT2and Wavele2based Image Coding [J]. IEEE Trans. Circuits and Systems for Video Technology, 1999, 9(5): 692- 695.

[11] 卓力, 沈兰荪, 等. 基于 DCT 变换的渐进式图像编码方法 [J]. 电子学报, 2002, 30(12A): 2105- 2107.

[12] SHEN Lan2sun, WANG Kong2qiao, Xing2xin. Automatic human face lo2cation and tracing in a complex background [J]. Chinese Journal of Electronics, 2000, 9(1): 65- 69.

[13] 邢昕, 汪孔桥, 沈兰荪. 基于器官跟踪的人脸实时跟踪方法 [J]. 电子学报, 2000, 28(6): 29- 31.

作者简介:

卓　力　女, 1971 年生于江苏徐州, 1992 年和 1998 年分别获得电子科技大学和东南大学学士和硕士学位, 现为北京工业大学讲师、博士生, 近年来发表论文多篇, 主要研究方向为图像/视频编码、无线视频传输等.

沈兰荪　男, 1938 年生于江苏苏州, 北京工业大学教授、博士生导师, 主要研究兴趣为图像与视频信号的压缩编码、处理与传输等, 著有/图像编码与异步传输0、/视频编码与低速率传输0等多部专著.