

一种基于小波域规则三角形网格运动补偿的可扩展视频编码方法

房 胜¹⁾ 梁永全¹⁾ 钟玉琢²⁾

¹⁾(山东科技大学计算机科学与技术系, 青岛 266510) ²⁾(清华大学计算机科学与技术系, 北京 100084)

摘 要 可扩展视频编码是指已编码码流在解码端能够进行部分解码,且重构图像的质量与接收到的信息量成比例变化,同时空域分辨率和时域分辨率可以根据解码端的要求改变。为了提供小波编码视频码流的高度可扩展性,提出了一种基于小波域规则三角形网格运动补偿的可扩展视频编码方法。该方法首先对参考帧使用冗余离散小波变换(RDWT)进行分解,并在 RDWT 小波域使用网格运动模型进行运动补偿,以提高运动补偿效率;其次在编码阶段则采用基于子带扫描顺序的改进 SPIHT 技术来保障码流的可扩展特性。实验表明,该方法在获得较高编码效率的同时,还提供了码流的高度可扩展特性。

关键词 可扩展性 冗余离散小波变换 网格运动补偿 SPIHT

中图分类号: TN919.81 **文献标识码:** A **文章编号:** 1006-8961(2007)01-0021-06

Highly Scalable Video Codec With Wavelet Domain Motion Compensation Based on Regular Triangular Mesh

FANG Sheng¹⁾, LIANG Yong-quan¹⁾, ZHONG Yu-zhuo²⁾

¹⁾(Department of Computer Science and Technology, Shandong University of Science and Technology, Qingdao 266510)

²⁾(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

Abstract Video scalability is referred to that the encoded bitstream can be decoded partly and the reconstruction video quality is proportional to the amount of received information, and the spatial and temporal resolutions can be changed according to the decoder property. This paper presents a highly scalable video codec with wavelet domain motion compensation based on regular triangular mesh, which decomposes the reconstructed reference frame by RDWT, carries out motion compensation in wavelet domain by mesh-based motion model, so that the efficiency of motion compensation is improved. In order to keep not only PSNR scalability but also spatial and temporal resolution scalability, an amended SPIHT method with subbands-scan order is proposed. The experimental results prove that the codec can produce encoded stream with good PSNR and highly scalable functions.

Keywords scalabilities, redundant discrete wavelet transform (RDWT), mesh-based motion compensation, set partitioning in hierarchical trees(SPIHT)

1 引言

在基于小波的视频编码技术中,运动估计/补偿(motion estimation/motion compensation, ME/MC)技

术是提高编码效率的关键技术。依所在位置划分目前主要有以下两大类运动补偿方法:一类是空域运动补偿方法^[1~4],另一类是小波域运动补偿方法^[5~9],但是从目前的技术看,空域 MC 方法对小波编码技术在运动补偿效率、可扩展性以及系数结构

基金项目:山东省自然科学基金项目(Y2002G07);山东省教育厅计划项目(J05C10)

收稿日期:2005-03-15;**改回日期:**2005-11-30

第一作者简介:房 胜(1971~),男。2002年清华大学计算机科学与技术系毕业并获工学硕士学位,现为山东科技大学副教授。主要研究方向为视频编解码、多媒体通信、无线 ad hoc 网络等。E-mail: fangsheng@tsinghua.org.cn

方面形成了制约^[10]。文献[6]提出了基于小波域的多分辨率运动估计算法,文献[7]提出了基于区域分割的小波域多分辨率运动估计算法,但是由于DWT(discrete wavelet transform)不具有空间平移不变性,因此无法在小波域内精确界定相邻帧间的运动,从而导致了大量匹配误差。文献[8]提出了LBS(low-band-shift)方法,用来进行ME/MC,虽取得了较好的编码效果,但是由于使用的小波块结构(wavelet block)^[11]来自于各个子带,因此码流不具备分辨率扩展特性。文献[5]在文献[8]的基础上提出了带内预测(in-band prediction)方法,同文献[8]中方法的差别在于它是以各子带中的块为匹配单位,和按子带顺序进行ME/MC,从而形成了全扩展码流。在具体实现时,传统的ME/MC方法在ME过程中大多使用基于块的运动补偿技术^[1-3,5-8]。这种技术假设所有块的运动都是平移运动,但是事实上图像序列中包含多种运动方式,如旋转、收缩和扩张等,这就限制了运动补偿的效率,而且从本质上来讲,基于块的运动补偿和基于整数的小波变换是不匹配的。文献[4]提出了3维小波编码系统中的可变形网格MCTF(motion compensation temporal filter)方法,即通过在空域构建规则形状的三角形网格提高多种运动模式情况下的运动补偿效率。文献[9]提出了一种基于小波域的任意三角网格运动补偿方法,但是如果要实现码流的可扩展性,则必须传递网格节点的位置,这样就增加了码流的额外负担。

本文提出一种基于小波域规则三角形网格运动补偿的可扩展视频编码方法。该方法的ME/MC过程是在冗余离散小波变换(redundant discrete wavelet transform, RDWT)的小波域中进行,并采用规则的三角形网格模型和利用RDWT域的空间位置不变特性

以及三角形网格模型对多种运动的适应性来提高运动补偿的效率。在编码时,SPIHT(set partitioning in hierarchical trees)技术^[2]虽然保障了码流的质量可扩展性,但是却破坏了小波系数的空间结构,为了保障码流的质量以及空域和时域可扩展性,本文提出了一种改进的子带扫描顺序SS-SPIHT技术(subband scan SPIHT),即通过加入一个子带标志,使小波系数按子带顺序进行扫描和编码,以便在保持质量可扩展性的同时,获得空域可扩展性。

2 基于小波域网格运动补偿的可扩展视频编码系统结构

本文提出的基于小波域规则三角形网格运动补偿的可扩展视频编码系统框架结构如图1所示。整个系统采用混合编码结构,主要由编码模块,隐含的解码模块和运动补偿模块3部分组成。整个系统的编码流程如下:当视频帧输入后,则首先判断是帧内编码还是帧间编码,如果是帧内编码,则直接进行DWT变换、SS-SPIHT编码和算术编码;然后进行SS-SPIHT解码以及DWT逆变换,再对得到的空域重构帧进行RDWT变换,就形成了小波域网格运动补偿方法中的“参考帧”,即RDWT各级子带;最后在小波域进行三角网格节点的选取和网格创建工作。如果是帧间编码,则首先对输入的空域图像进行RDWT变换,变换完成后得到的RDWT各级子带,即是小波域网格运动补偿方法中的“当前帧”;然后对“参考帧”中的网格节点进行前向追踪,也就是在“当前帧”中通过进行运动估计来获得网格节点的偏移量,也就是运动矢量。RDWT小波域的“当前帧”和“参考帧”的网格拓扑结构相同,而且由

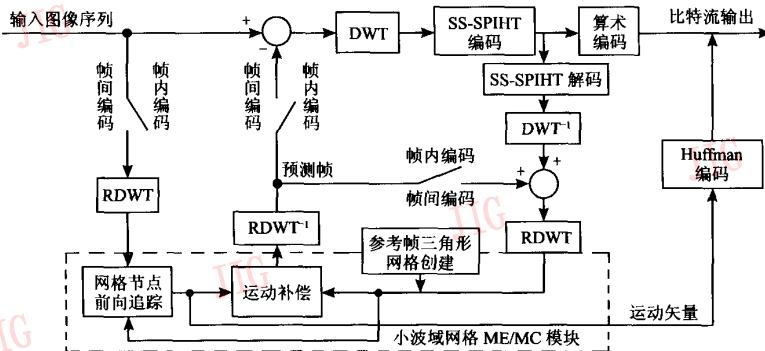


图1 基于小波域网格运动补偿的可扩展视频编码方法的框架结构

Fig. 1 The proposed framework of scalable video codec with wavelet domain mesh motion compensation

于网格节点的运动矢量建立了一个从“参考帧”到“当前帧”的节点间的对应关系,因此通过这些对应关系就可以实现一系列的从“参考帧”到“当前帧”的映射,映射函数的输出即是“当前帧”的RDWT小波域的“预测帧”;最后把该预测值通过逆RDWT变换到空域,就形成了空域图像的预测。这样从原始图像中减去该预测值即得到残差帧,虽然对残差帧的后续处理同帧内编码基本是相同的,但其主要的差别是解码过程中的DWT逆变换完成后,其得到的结果需要与空域图像的预测帧进行相加,才能得到完整的重构图像。

2.1 RDWT变换的实现

小波域网格运动补偿技术的一个重要基础是RDWT的空间平移不变性,而且RDWT可以看作是连续小波变换的一种近似。在滤波器组实现方式中,RDWT同实现临界采样DWT的Mallat算法相比,其主要差别在于:Mallat算法在每级分解后都要进行下采样,而RDWT算法则去掉了下采样环节,从而得到的是过采样的输出系数。RDWT变换有多种实现方

式,如LBS(low-band-shift)方法^[8]以及“多孔算法”^[12]。对1维信号,“多孔算法”的实现如下列公式所示。RDWT每一级的分解过程中,分析过程为

$$a_{j+1}(k) = a_j(k) * g_j(k) \quad d_{j+1}(k) = a_j(k) * h_j(k) \quad (1)$$

而滤波器组则需要上采样(\uparrow 表示上采样),以适应增长的数据,对于分解级别 j 来说,

$$h_{j+1}(k) = h_j(k) \uparrow 2 \quad g_{j+1}(k) = g_j(k) \uparrow 2 \quad (2)$$

而重构过程为

$$a_j(k) = \frac{1}{2}(a_{j+1}(k) * g_{j+1}(k) + d_{j+1}(k) * h_{j+1}(k)) \quad (3)$$

式(1)、式(2)和式(3)还可以扩展到2维的情况。2维RDWT变换可得到 $3 \times L + 1$ 个子带(L 为分解级),这些子带与原始图像具有相同的分辨率(如图2所示),并且由于变换的线性不变特点,各子带在相同位置上的系数来自原始图像中的同一个像素,这就为三角形网格的创建以及进行ME/MC提供了条件。

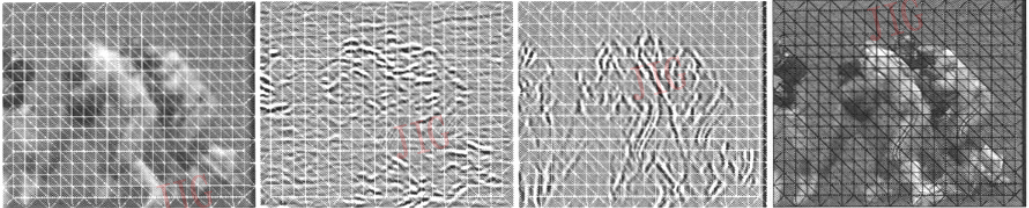


图2 一个规则三角形网格示例,覆盖Football序列重构第1帧及其RDWT子带

Fig. 2 Example of Regular triangular mesh in first reconstructed frame of Football and it's RDWT subbands

2.2 三角形规则网格的创建

三角形网格在运动补偿的使用中,可以分为规则三角形和不规则三角形网格两大类。由于规则三角形网格在构建时,其节点位置是确定的,因此不需要传递节点的位置;而不规则三角形网格节点应尽量放置在物体的边缘及纹理等特征点上^[13],因为这些位置往往是图像运动场发生分化的关键点,但是由于这样就必须在码流中传递节点的位置信息,因此就增加了码流的额外负担,尤其不利于低码率下的编码,而且在可扩展码流中,解码的码流往往与编码时的码流不同,即使使用相同的构造算法得到的网格结构也会不同。考虑到码流扩展的需要,本文采用规则的三角形网格。

考虑到码流可扩展性的要求,尤其是空域可扩展性的要求,图像内部的网格节点应以16的整倍数

位置进行选取,图像边界的节点以32的整倍数位置进行选取。在节点选择完毕后,再使用Delaunay方法^[11]来构造三角形网格,这样就可将构造的网格拓扑结构应用到各分解级的各个子带中(如图2所示),图中从左向右依次是RDWT3级分解的基带、水平子带、垂直子带以及重构图像。从图中可以看出,构建的三角形网格对RDWT的各级子带和空域图像的适应性是相同的。

2.3 运动估计和运动补偿:网格节点的前向追踪过程以及网格映射

运动补偿技术中运动估计的准确度是决定运动补偿效率的关键。本文采用三角形网格节点的前向追踪方式来进行运动估计。由于运动估计是发生在RDWT域,因此需要综合各个子带的信息作为判据。运动矢量(motion vectors)由式(4)确定,式(4)中, w

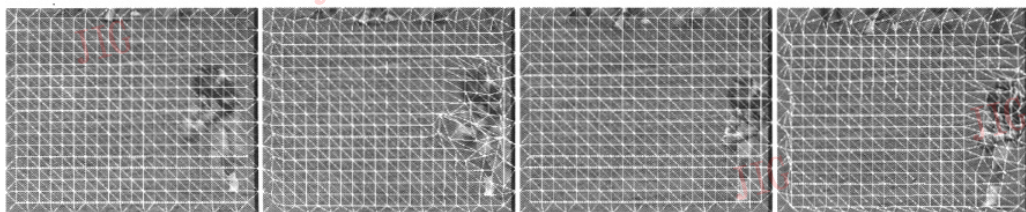


图3 网格节点的前向追踪,即运动估计过程

Fig. 3 The forward tracking of mesh nodes, equals the function of motion estimation

是匹配过程中的搜索范围。

$$MV_s = \{u, v \mid \min MAD(x, y, u, v), -w \leq u, v \leq w\} \quad (4)$$

其中,平均绝对差(mean absolute difference, MAD)的计算式为

$$MAD(x, y, u, v) = \frac{1}{B^2} \sum_{m=1}^B \sum_{n=1}^B AD(x+m, y+n, u, v) \quad (5)$$

其中,绝对差(absolute difference, AD)为

$$AD(x, y, u, v) = \sum_{l=1}^L \left\{ \frac{1}{2^l} \left[|HL_l^{cur}(x, y) - HL_l^{ref}(x+u, y+v)| + |LH_l^{cur}(x, y) - LH_l^{ref}(x+u, y+v)| + |HH_l^{cur}(x, y) - HH_l^{ref}(x+u, y+v)| \right] + \frac{1}{2^l} |LL_l^{cur}(x, y) - LL_l^{ref}(x+u, y+v)| \right\} \quad (6)$$

B 为以节点为中心的块的尺寸, LL 、 LH 、 HL 和 HH 分别代表基带、水平、垂直和对角子带, L 表示分解级数, l 表示当前分解级, cur 表示当前帧, ref 表示参考帧。

在当前帧和参考帧的 RDWT 域的各子带之间, 由于节点的运动矢量建立了一个从参考帧到当前帧的节点间的对应关系, 因此通过这些对应关系就可以实现一系列的从参考帧到当前帧的映射, 也即运动补偿。映射是以网格中的每个三角形片为单位, 而映射函数则采用仿射模型, 该方法的具体实现可参考文献[14], 本文不再进行描述。需要考虑的问题是, 当图像中的运动矢量较大时, 由于三角形网格的拓扑结构会发生反转等蜕化现象, 因此会影响映射的效果, 最终会降低运动补偿的效率。为了克服此困难, 又兼顾到计算复杂度, 本文方法对运动剧烈的图像序列, 如 Football 序列, 采用每两帧重构一次

三角形网格的方法。

图3显示了Football序列中连续4帧的网格拓扑结构及其变化, 图中三角形网格绘制在重构图像上(码率1.0bpp)。图中从左至右依次为42, 43, 44和45帧, 在这几帧中由于图像中的对象较少, 因此可以比较清楚地分析三角形网格结构的变化。从图像上方观众的腿部位置的变化可以看出, 相机有强烈的平移运动, 对象(队员和球)也在剧烈运动。通过同42帧和44帧进行对比可以看出, 43帧和45帧上方的三角形网格结构的变化反映出了相机运动造成的场景变化, 而覆盖着对象区域的网格也循着对象的运动轨迹进行变化(平移, 收缩或者旋转), 如在44帧和45帧中, 队员右腿在摆动的过程中, 网格节点沿对象的边缘及重要的纹理进行移动, 而且网格的覆盖区域与对象的运动轨迹是对应的。从以上分析可以看出, 在RDWT小波域构造规则三角形网格的方法不仅能够充分地反映图像中物体的运动轨迹, 而且有利于提高运动补偿的效率, 试验结果也证明了此点。

3 码流可扩展性的实现

SPIHT编码技术虽然是小波编码技术中效率较高的一种方案, 其产生的码流虽具有质量可扩展性, 但是编码过程中同一子带中相邻像素的空间关系却被丢失掉了, 也即分辨率扩展的特性丢失了。本文提出的改进的子带顺序扫描SPIHT编码方法, 即SS-SPIHT方法虽使用和SPIHT算法相同的原则进行系数的重要性以及集合重要性的分类, 但是在每一个小波系数结构中再加入一个标志, 该标志表明该系数属于DWT变换的哪一个分解级, 这样在进入队列时, 队列中的系数就按子带顺序进行排列。解码时可使用相同的结构, 这样就在保证码流的质量可扩展性的同时, 还保障了小波系数结构的空域关系。码流的编码顺序参见图4。编码时, 首先按

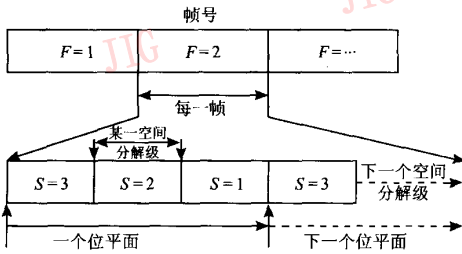


图 4 编码过程中的系数扫描顺序

Fig. 4 The scan order of coefficients in encoding

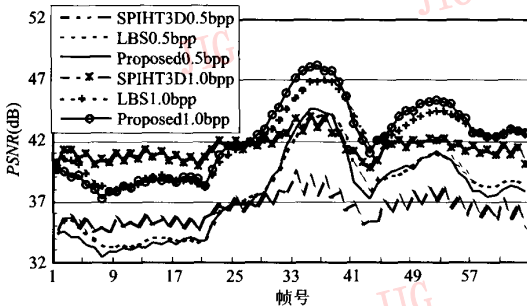
帧进行编码,而在每帧中则按位平面(bit plane)进行编码,在位平面中则按各分解级进行编码。文献[3]中提出了一种 SPIHT 的改进技术——全扩展零树编码方法(fully scalable zerotree coding, FSZ)。这种方法是在编码的小波系数结构中加入一个标志,表示系数要执行的操作是哪一种,即集合重要和/或者像素重要,实际执行时,进入队列的操作则变为仅是改变系数的标志,也即最后编码时,小波系数仍然是按子带顺序进行的,这和本文提出的方法思路一致,只是具体实现的方法有所差别。

4 实验结果

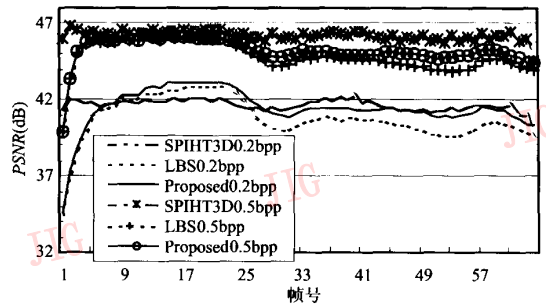
为验证本文方法的效果,采用 Footaball 序列和 Mother & Daughter 序列,对本文提出的基于小波域网格运动补偿的可扩展视频编码方法进行了测试实验,并同 LBS 方法^[8]和 3D-SPIHT 方法^[2]进行了比较。测试时,除第 1 帧采用帧内编码外,其余各帧均采用帧间编码方式。在运动补偿过程中,使用的 RDWT 变换以及对图像残差进行的 DWT 变换,均采用 CDF9-7 小波来实现 3 级分解。为了比较更客

观,LBS 方法也采用此种设置,3D-SPIHT 方法中时域的分解级别是 3 级,即 GOF (group of frames) 为 8 帧。测试序列分别选择运动剧烈的 Football 和运动不剧烈的 Mother & Daughter 序列,编码格式为 CIF,序列长度为 64 帧。编码 Football 序列采用 0.5bpp 和 1.0bpp 压缩码率, Mother & Daughter 序列采用 0.2bpp 和 0.5bpp 压缩码率。

测试 Football 序列时,式(4)中的搜索范围 w 为 11,式(5)中的块尺寸 B 为 15。由于 Football 序列中存在大量的快速移动对象以及相机运动,因此在三角形网格的追踪过程中采用每 2 帧重置一次网格结构的方法;在 Mother & Daughter 序列中,由于对象运动以及场景变换都不是很剧烈,式(4)中的搜索范围 w 为 9,式(5)中的块尺寸 B 为 15,每 10 帧重构一次三角形网格。试验结果如图 5 所示,其中 Proposed 是指本文提出的方法。实验结果表明,本文提出的方法在运动剧烈的 Football 序列中,与 SPIHT3D 方法相比,在 0.5bpp 码率下,解码图像的峰值信噪比(peak signal noise ratio, PSNR)高出约 1.2dB,在 1.0bpp 码率下,解码图像的 PSNR 高出约 0.75dB;与 LBS 方法相比,在 0.5bpp 码率下,解码图像的 PSNR 比 LBS 方法约低 0.25dB,在 1.0bpp 码率下,解码图像的 PSNR 比 LBS 方法高出约 0.3dB。而在运动不剧烈的 Mother & Daughter 序列中,与 3D-SPIHT 方法相比,在 0.2bpp 码率下,解码图像的 PSNR 约低 0.2dB,在 0.5bpp 低码率下,解码图像的 PSNR 约低 0.95dB;与 LBS 方法相比,在 0.2bpp 码率下,解码图像的 PSNR 约高 0.6dB,在 0.5bpp 码率下,解码图像的 PSNR 约高 0.4dB。以上 PSNR 数值均指图像序列 64 帧的平均数值。从实验结果可以看出,本文提出的基于小波域规则三



(a) Football 图像序列



(b) Mother & Daughter 图像序列

图 5 用本文提出的方法和 LBS, SPIHT3D 方法的解码图像序列的峰值信噪比比较

Fig. 5 The PSNR comparison between image sequences decoded by proposed method and LBS, SPIHT 3D

角形网格运动补偿的可扩展视频编码方法对多种运动情况下的图像序列适应性较好,这与先前的分析是一致的。而在运动比较剧烈的图像序列中,如 Football 序列,其编码质量有所下降,尤其是在低码率下,这主要是由于基于三角网格的运动模型,对于网格节点的位置更加敏感的缘故,由于在 Football 序列中网格节点的偏移量较大,因此经常出现三角形片的几何形状发生反转等蜕化现象。在程序实现时,考虑到计算的复杂度,由于本文采用略过蜕化三角形片的方法,因此造成了 Football 序列中的运动补偿效率有所下降。在运动程度较低的 Mother & Daughter 序列中,由于此类图像中的时域相关性非常强,因此 3 维小波编码中的时域滤波能够较好地发挥作用,这也从另一个方面证明了运动补偿对视频编码的重要性。对于低空域分辨率下的解码图像,由于作为参考对象的原始图像不能确定,故本文没有给出相应的峰值信噪比数值。

5 结 论

本文提出基于小波域网格运动补偿的可扩展视频编码方法,结合小波域运动补偿的空间不变特性和三角形网格对多种运动模型的良好适应性,则不仅可提高运动补偿的效率,并且通过改进 SPIHT 编码技术,还保障了码流的空域可扩展特性,从而在获得较好编码效率的同时,还为码流提供了质量、空域和时域等的可扩展特性。这对于码流在异构网络中的传输是非常重要的。

冗余离散变换小波由于其变换过程中的空间不变特性,从而在视频编码技术和图像编码技术中都受到了较高的重视^[10],还可以进一步对 RDWT 域运动补偿特性进行研究和利用,如在运动估计的匹配算法中,就可以集成重叠运动补偿方法;同时提出的小波域的网格运动补偿方法还可以扩展到 3 维小波编码系统中,这将是今后工作的重点。

参考文献 (References)

- 1 Taubman D. Directionality and Scalability in Image and Video Compression[D]. University of California at Berkeley, CA, USA, 1994.
- 2 Kim B J, Z Xiong, Pearlman W A. Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees(3D SPIHT)[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(8): 1374 ~ 1387.
- 3 Botreau V, Benetiere M, Felts B, et al. A fully scalable 3D subband video codec[A]. In: Proceedings of IEEE International Conference Image Processing'01[C]. Thessaloniki, Greece, 2001, 2: 1017 ~ 1020.
- 4 Secker A, Taubman D. Highly scalable video compression using a lifting-based 3d wavelet transform with deformable mesh motion compensation[A]. In: Proceedings of the International Conference on Image Processing[C]. Rochester, New York, USA, 2002, 3: 747 ~ 752.
- 5 Andreopoulos Y, Munteanu A, Van Der Auwera, et al. Wavelet-based fully-scalable video coding with in-band prediction[A]. In: Proceedings 3rd IEEE Benelux Signal Processing Symposium (SPS-2002)[C], Leuven, Belgium, 2002: 217 ~ 220.
- 6 Zhang Y-Q, Zafar S. Motion compensated wavelet transform coding for color video compression[J]. IEEE Transactions on Circuits and System for Video Technology, 1992, 2(3): 285 ~ 296.
- 7 Liu Hong-mei, Huang Jin-ke, Xiao Zi-mei. A region-segmentation-based multiresolution motion compensation wavelet video coding algorithm[J]. Journal of Image and Graphics, 2002, 7A(3): 292 ~ 296. [刘红梅, 黄金科, 肖自美等. 基于区域分割的小波分解运动补偿编码算法[J]. 中国图象图形学报, 2002, 7A(3): 292 ~ 296.]
- 8 Park H W, Kim H S. Motion estimation using low-band-shift method for wavelet-based moving-picture coding[J]. IEEE Transactions on Image Processing, 2000, 9(4): 577 ~ 587.
- 9 Cui S, Wang Y, Fowler J E. Mesh-based motion estimation and compensation in the wavelet domain using a redundant transform[A]. In: Proceedings of the International Conference on Image Processing [C]. Rochester, USA, 2002, 1: 693 ~ 696.
- 10 Ohm J-R, Schaar M van der, Woods J W. Interframe wavelet coding-motion picture representation for universal scalability [J]. Image Communication Journal, Special Issue on Digital Cinema, 2004, 19(9): 877 ~ 908.
- 11 Shewchuk J R. Triangle: engineering a 2D quality mesh generator and delaunay triangulator [A]. In: (Ming C. Lin and Dinesh Manocha, editors) Applied Computational Geometry: Towards Geometric Engineering, Lecture Notes in Computer Science[C], Berlin, Germany: SpringerVerlag, 1996: 203 ~ 222.
- 12 Yang Fu-sheng. Wavelet Transform in Engineering Analysis and Application[M]. Beijing: Science Publisher, 1999. [杨福生. 小波变换的工程分析与应用[M]. 北京: 科学出版社, 1999.]
- 13 Wang Y, Lee O. The use of deformable mesh structure for video coding part II: The analysis problem-mesh generation, nodal displacement estimation, and element classification [J]. IEEE Transactions On Circuits and Systems for Video Technology, 1996, 6(6): 647 ~ 659.
- 14 Wang Y, Lee O. The use of deformable mesh structure for video coding part I: The synthesis problem-mesh-based function approximation and mapping[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(6): 636 ~ 646.