

## 基于多分辨率二叉树的运动分割及 交叠块的立体视频帧估计算法

骆艳 张兆扬

(上海大学通信与信息工程学院, 上海 200072)

**摘要** 为了对立体视频序列进行有效的编解码, 以获得较好的重建图象, 提出了一种基于立体视频序列的帧估计及内插方法. 该方法就是在立体视频序列编码中, 左图象流按 MPEG-1 / I 的视频编码标准进行编码; 而对右图象流, 只对参考帧(I 帧及 P 帧)进行编码, 右图象流的 B 帧不进行编码及传送, 而是在解码端, 根据各个参考帧通过帧重建及内插来获得. 可变块分割的方法是采用基于多分辨率二叉树分解的运动分割. 对右 B 帧的重建按照匹配块和失配块分别处理的方法, 各个块的位置和内容则根据右 B 帧中的块与右 I 帧、P 帧之间或与左 B 帧之间相关性大小来确定. 另外, 还提出了一种新的可避免误匹配和对显露区进行正确填充, 并针对交叠块的帧估计和内插方法. 仿真实验表明, 该图象重建方法与文献[1]所提出的基于固定块的方法相比, 能够使得重建图象的峰值信噪比(PSNR)增加约 1.25dB, 并且图象的主观视觉质量评价也明显要好.

**关键词** 立体视频序列编码 二叉树分割 多分辨率 帧估计和内插技术

中图法分类号: TN919.81 文献标识码: A 文章编号: 1006-8961(2002)07-0716-07

## The Frame Estimation Algorithm of Stereo Video Sequences Based on Multi-Resolution Quadtree-Based Motion Segmentation and Overlapped Blocks

LUO Yan, ZHANG Zhao-yang

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

**Abstract** In order to code and decode the stereo video sequences effectually and acquire excellent reconstructed images, a frame estimation and interpolation technique of stereo video sequences is presented in this paper. In stereo sequence coding, the left image stream is coded according to the MPEG-1 / I video coding standard. In the right image stream, only reference frames (I and P) are coded, the B frames in the right image stream are neither coded nor transmitted, they are acquired from the reference frames with the frame reconstruction and interpolation technique at the decoding end. The variable block segmentation is based on a multi-resolution quadtree-based motion segmentation. This paper also proposes a novel method to process the reconstruction of right B frame as matched-block and unmatched-block separately, the positions and content of these blocks are decided on the correlation between the blocks in right B frames and right I, P frames or left B frames. A novel frame reconstruction and interpolation technique for overlapped blocks to avoid false matching and to fill the exposed regions correctly is also proposed in this paper. The simulate experiment results show that the average PSNR is about 1.25dB higher than the method mentioned in the reference [1], which is based on fixed blocks, and the reconstructed frames have a much better presentation on viewing perception.

**Keywords** Stereo sequence coding, Quadtree-based segmentation, Multi-resolution, Frame estimation and interpolation

基金项目: 国家自然科学基金项目(69972027)

收稿日期: 2001-06-20; 改回日期: 2001-12-10

## 0 引言

立体电视很可能是继 HDTV 之后的下一代电视,其立体图象显示给人们提供了一个从左右两幅图象中获得真实世界相对深度感的直接而简单的方法。由于这种相对深度信息在航空图象立体观察,远程控制,医学和微观研究中是至关重要的,因此 MPEG-4 中也支持立体视频和多视点编码功能。但影响立体视频发展的主要障碍是需要传输和储存的信号带宽比单视信号要多一倍,如果每个视的视频信号都进行单独编码的话,这是不可接受,也是不实际的。由此可见,研究如何有效地传输和存储立体或多视视频数据,提高压缩率是十分重要的。

根据人眼在一只眼感觉图象清晰,而另一只眼感觉图象稍模糊时,其综合感觉仍然是清晰的,并且具有三维的视觉特点,因此可将其中对应于一只眼看到的视频信号,以高分辨率等级,按传统方法编码,称为主视流,而对应于另一只眼看到的视频信号可通过对主视的视差补偿预测(DCP)来获得,并以较小带宽的图象流传送,称为辅视流。这样在解码端解码后,其所得的分辨率不一的两个图象,经双目融合后即可刺激出立体视觉<sup>[2]</sup>。在本文中规定对应左眼的图象流为主视流,对应右眼的图象流则为辅视流。

对上述的主视流可用类似 MPEG 的帧结构来进行描述。辅视流中的 I 帧可由主视流的 I 帧经视差补偿预测(DCP)来得到;辅视流中的 P 帧可以通过辅视流的 I(或 P)帧运动补偿预测(MCP)和主视流中相同时刻 P 帧的视差补偿预测(DCP)获得;依此类推,辅视流中的 B 帧可以通过辅视流的双向 MCP 和与主视流对应的 B 帧的 DCP 来进行重建<sup>[1-3]</sup>。更进一步地说,不需要进行编码传输,可直接在解码端通过帧估计和内插来进行重建。因 B 帧数目大于 I 帧和 P 帧之和,这将显著降低立体视频序列的传输比特率。上述的关系可由图 1 表示。图 1 中的下标 M(Main Stream)表示主视流;A(Auxiliary Stream)表示辅视流。

问题的关键是怎样从参考帧(右 I 帧、右 P 帧、和同一时刻的左 B 帧)中获得对相应的右 B 帧进行准确估计的所有信息。对此,文献[1]曾提出了一种基于固定块的方法。该方法是首先在右 I 帧和 P 帧中进行块运动匹配,然后将搜索到的运动矢量,按 B 帧到 I 帧的帧距离进行比例分配,从而得到新的运动矢量,

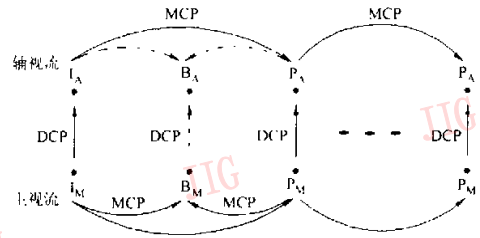


图 1 立体序列的编解码方法

DCP: 视差补偿预测; MCP: 运动补偿预测  
(虚线表示右 B 帧在编码端不编码及传送,而在解码端通过帧重建及内插方法获得)

再将该块按新的运动矢量映射到相应的图象位置,该块的图象内容则根据参考帧组合的加权和而得到。

## 1 基于固定块的右 B 帧重建及内插方法所存在的缺陷

文献[1]提出的基于固定块的重建右 B 帧及内插方法存在以下缺陷:

(1) 该方法采用大小为  $16 \times 16$  的固定块,由于宏块相对图象某些细节的尺寸较大,且在一些区域的宏块内,各像素的运动性质并不一致,同时又忽略了图象中运动对象的亮度变化,因此在 MAD 匹配准则下,选择匹配块时,将产生一种“多数”效应,即求得的匹配宏块,仅表现出原始宏块内占多数像素区域的运动性质,从而会导致某些运动估值的区域性失败。

(2) 由于该方法对所有块的图象内容是直接根据参考帧组合的加权和得到,而没有考虑各个参考帧之间的物体运动和视差遮挡,因而在物体边缘处存在着遮挡和暴露区域。这种加权和的方法类似于边缘平滑滤波器,它可造成在物体边缘处形成明显的“模糊”虚像。在运动区域(尤其是出现背景遮挡/暴露,或块中的运动物体出现形变时),按这 3 个参考块的加权平均所获得的重建块,会在物体边界处出现模糊的重影(如图 2 所示)。

(3) 这种基于固定块的右 B 帧重建方法是将各个子块,按运动矢量大小分为静止块和运动块进行处理,而文献[1]认为在运动区域,对于运动块来说,右 B 帧与左 B 帧的相关性要比与右 I 帧、P 帧的相关性大。这意味着,左 B 帧所提供的运动物体的信息要比另两个参考帧多,因此文献[1]在加权过程中,对左 B 帧的内容给予较大权重。然而,如果帧之间存在摄像机运动的话,则图象中的场景会发生全



图2 当参考帧组合中某块的内容出现局部形变时,基于固定块方法得到的重建块示意图  
图中字母  $B$  表示图象块(block), (下标  $L$  表示左;  $R$  表示右;  $I, B, P$  分别代表  $I$  帧,  $B$  帧,  $P$  帧)

局运动,这样各个子块均为运动块,若对每个运动块均采用视差搜索的方法,则将增加计算量,而且,根据立体摄像机几何特性,当左右摄像机满足平行光轴条件时,则同一时刻的场景在左右摄像机平面上所得到的两幅图象,就可以看作是同一台摄像机对一幅静止场景在相距一定平行距离上分别拍摄到的图象.为了能保留足够的深度信息,以便在显示屏上呈现出立体感,则必须使左右摄像机之间的距离不能太近.另外,除少数情况(摄像机快速位移)外,占图象绝大部分的静止区域,在同一视内的  $I$  帧、 $P$  帧之间,由于摄像机位移造成的视差和遮挡要比左右立体视图象少得多.尤其在拍摄过程中,即使摄像机缓慢移动或摄像机没有位移的情况下,同一个视内的视频序列,其各帧之间的相关性也要远远大于同一时刻左右视间图象的相关性.

由于以上原因,左  $B$  帧与右  $B$  帧中宏块的相关性不一定比与右  $I$  帧、右  $P$  帧的相关性大,且由于运动遮挡,故右  $I$  帧中的块不一定能在右  $P$  帧中找到合适匹配块(失配块),并由此造成误匹配,由此可见,对所有运动块图象内容进行加权和方法在存在摄像机运动时,往往造成解码端,解码速度的放慢以及伪轮廓和误匹配增加,最后将导致信噪比的下降.

(4) 通常在对交叠块(因运动造成某些块之间的部分交叠)进行估计时,将造成一些区域、因缺少匹配覆盖而形成“空洞”,而一些区域则出现多重匹配,因此,需要进行如下两种处理:

- ①在多重匹配中选择正确匹配;
- ②对大面积未被覆盖的暴露区域进行正确填充.

在文献[1]中,对第①种处理未作讨论,而对第②项的处理,则是利用该区域与其周围块之间的相关特性,首先将这些块的视差矢量作为候选矢量,并将其中找到的发生次数最多的视差矢量作为最佳矢量;然后根据视差矢量找到该区域在左  $B$  帧中的位置,并从视差矢量来得到左  $B$  帧相应的估计区域,再将该区域填充到该位置处,从而完成对遮挡和失配区域的处理,但这样的处理会造成如图3所示的误填充.如图3(b)中,白条处即为未被覆盖,属于暴露出来的背景区域,若该区域正好是位于两个具有较大视差的区域之间的话,文献[1]提出的填充方法就会造成将左  $B$  帧中具有较大视差区域的内容填入该未被覆盖区域(图3(c)).另外,由于未被覆盖区域形状的不规则性,因而对各个邻域块进行视差估计将造成计算量的增加和重建图象速度下降,这违背了在解码端快速重建图象的要求.



(a) 原始右  $B$  帧图象 (b) 存在显露区域的重建帧图象 (c) 误填充后的右  $B$  帧图象

图3 造成误填充的说明

## 2 基于多分辨率二叉树分解的运动分割及交叠块帧估计算法

### 2.1 基于运动和亮度信息的多分辨率可变尺寸块

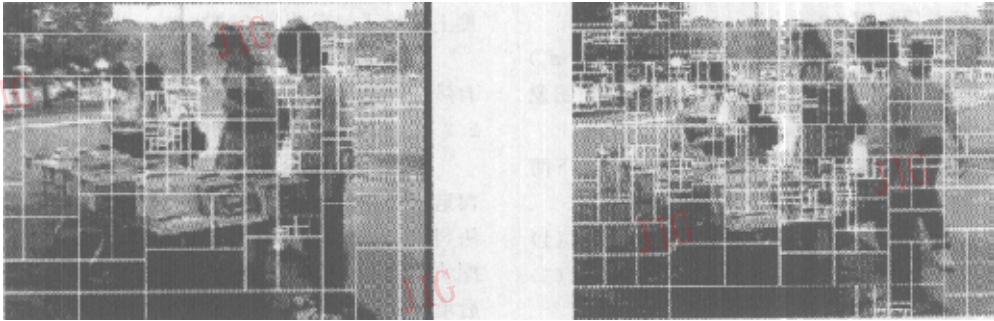
#### 分割

在块分割上,本文采用文献[4]、[5]中提出的多分辨率二叉树可变块分割方法进行分割,并作了改进.文献[4]、[5]的方法是针对静止立体图象对的视差信息进行分割的,分割时,同时利用了视差和

图象亮度信息,但由于视差的全局分布特性以及位于物体边缘区域的视差遮挡,使得分割的子块数较多,从而造成后续的帧估计重建速度下降.本文则是利用辅视流序列中的右 I 帧和 P 帧间的运动和边缘亮度信息来进行分割.分割时,首先对 P 帧和 I 帧按照从低到高的分辨率等级进行运动匹配搜索,并将在低分辨率级所获得的运动矢量作为更高分辨率级块匹配搜索的初值;然后当探测到的搜索块与原始块的 MAD 值大于该级的运动匹配阈值时,即进行块分割.分割分别按行和列进行,分割点分别位于块中水平方向(行)或垂直方向(列)的亮度不连续点上.这样分割就发生在与尖锐运动不连续的地方最接近的位置(如物体的边缘).由于在同一视内,帧与帧之间因物体运动而产生的遮挡/暴露区域只是局部存在的,即使是存在摄像机运动的情况下,帧与

帧之间仍存在着大量的匹配区域,因此选择在同一视内的 I 帧、P 帧图象进行运动估计和分割,就可减少分割块的数目,这样不仅可大大减少后面的帧估计和“空洞”填充的计算量,且可提高合成右 B 帧的速度.用改进的分割算法对 Booksale(320×240)视频序列第 40 帧和第 56 帧图象进行分割的结果如图 2(a)和(b)所示.

从图 2 可以看到,利用对占图象绝大部分的静止区域进行粗分割,而对位移帧差大的运动区域进行细分割的分割方法,可提高运动估计的精度.由于分割点位于物体边缘处,因而尽可能地减少了块运动估计中存在的块状可视干扰.这种分层的运动估计的方法不仅可使估计的抗噪性能逐渐加强,而且检测大的运动的能力也得到增强,从而使前节方法中的第 1 个缺陷得到较好克服.



(a) 在参考帧中存在大量静止区域时的分割结果(145 块)

(b) 在参考帧中运动区域较多时的分割结果(471 块)

图 4 基于运动的多分辨率分割结果(从 I 帧到 P 帧)

## 2.2 重建右 B 帧的算法

在完成对 P 帧的运动估计和分割后,将所有子块分为匹配块(包括静止区域和运动匹配区域)和失配块(运动遮挡和背景暴露区域)两种,并分别按其性质进行交叠块的估计和填充.具体算法如下:

令字母  $B$  代表图象中的任意一个子块,上标 I, P, B 分别为代表 I 帧, P 帧和 B 帧图象,而下标 L 和 R 表示左图象和右图象.

设  $B_k^I(x_k)$  为右 I 帧图象中的大小为  $h \times w$ , 中心在  $x_k$  的块,  $k$  表示该块的序号,指向右 P 帧图象的运动矢量为  $v_k$ , 则在右 P 帧图象中的匹配块为  $B_k^P(x_k + v_k)$ , 其平均绝对差为

$$D = \frac{1}{h \times w} \| B_k^I(x_k) - B_k^P(x_k + v_k) \| \quad (1)$$

设定一个匹配阈值为  $T$ ,

(1) 当  $D \leq T$  时, 定义为匹配块;

(2) 当  $D > T$  时, 定义为失配块.

### 2.2.1 对匹配块区域的处理

当  $D \leq T$  时, 意味着在右 P 帧图象中能找到合适的匹配块. 根据运动的连续性和帧到帧之间的时间相关性原理, 在右 B 帧图象中相应的运动匹配块的位置可由下式得到:

$$v_k = \frac{m}{m+n} v_k' \quad (2)$$

式中,  $m$  为 I 帧与 B 帧图象间的帧数,  $n$  为 B 帧与 P 帧图象间的帧数,  $v_k'$  为 I 帧到 B 帧图象的运动矢量估计.

在得到 B 帧图象的运动块位置  $(x_k - v_k')$  后, 可不作视差估计, 而直接通过右 I 帧, P 帧图象中相应参考块的平均来得到右 B 帧图象中的块  $B_k^B(x_k - v_k')$  的估计

$$B_k^B(x_k - v_k') = (B_k^I(x_k) + B_k^P(x_k + v_k)) / 2 \quad (3)$$

这样的帧估计和重建方法,由于避免了在右B帧图象中进行视差搜索,因而可显著减少计算量。

### 2.2.2 运动失配区域的处理

当  $D > T$  时,意味着该块位于I帧到P帧图象的运动遮挡区域或背景暴露区域。由于右I、右B、右P帧图象是不同时刻拍摄到的一个运动物体图象,而左B帧图象则是与右B帧同一时刻拍摄到的运动物体图象,仅透视投影位置稍有偏移,因此,这部分区域内容可以从左B帧图象中获取。

由于立体视频序列邻接帧视差矢量场和运动矢量场之间存在着高度的时间、空间相关性,且这种相关性类似“闭环”的关系<sup>[6]</sup>(见图5),因此按照图5运动与视差估计的关系,就可以从左B帧图象中获得这部分区域,其失配块的重建方法为:

(1) 对于右I帧图象中的  $B_R^I(x_k)$  块,在左I帧图象中进行视差估计,通过获得的视差矢量  $d_k$  找到左I帧图象的对应块  $B_L^I(x_k + d_k)$ ;

(2) 在左B帧图象中,对左I帧的  $B_L^I(x_k + d_k)$  块进行运动估计,得到运动矢量  $v_{kL}$  以及左B帧图象的对应块  $B_L^B(x_k + v_{kL} + d_k)$ ;

(3) 获得右B帧图象块的位置和内容有如下两种可选方法

$$\text{方法 ①} \quad B_R^B(x_k + v_{kL}) = B_R^I(x_k) \quad (4)$$

$$\text{方法 ②} \quad B_R^B(x_k + v_{kL}) = B_L^B(x_k + v_{kL} + d_k) \quad (5)$$

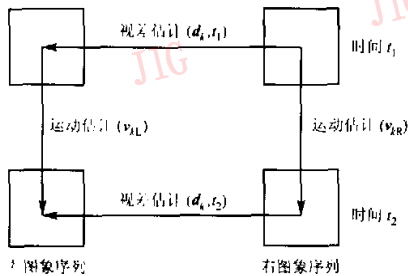


图5 立体运动合成

根据右B帧图象与右I帧和左B帧图象的相关性大小,分别选用方法①或②来获取右B帧图象块的位置和内容。在背景和无运动区域(在2.1节进行分割时,所获得的运动矢量  $v_{kL}$  小于一定的阈值),右B帧与右I帧图象的相关性要大于与左B帧图象的相关性,因此在此区域内可选用方法①;在运动区域( $v_{kL}$  大于该阈值),因右B帧与左B帧图象的相关性要比与右I帧和右P帧图象的相关性大,故在此区域内,可选用方法②。由于这样处理的好处是明显抑

制第1节方法第2个缺陷的“虚像”和“重影”,从而提高了重建图象的质量。

阈值  $T$  选取对右B帧图象的重建速度和结果会造成影响,考虑以下两个极端情况:

当  $T \rightarrow \infty$  时,则右P帧图象中所有的块都为匹配块,对这些块均按照对匹配块区域的处理方法进行。由式(2)、式(3)可见,这相当于单视序列中对B帧图象的重建和线性内插。这虽简单快速,但由于没有利用到立体视频的特点,在出现运动遮挡/背景暴露时,会出现第1节方法第2个缺陷的“虚像”和“重影”,从而降低了重建帧图象质量;

当时  $T \rightarrow 0$ ,则右P帧图象中的失配块数目增加,对这些块可按照对运动失配区域处理的方法进行处理,这虽然可明显抑制第1节方法第2个缺陷的“虚像”和“重影”,并可提高重建图象的质量,但是对每一个块分别进行运动或视差估计,则在一定程度上牺牲了解码端的重建速度。

因此, $T$ 的取值,经权衡、折衷,并考虑图象帧之间自然光照变化和噪声等因素,一般取在8~16之间。

### 2.2.3 避免误匹配的处理及对显露区的填充

为避免第1节方法第3个缺陷中,因多重匹配而造成的可能误匹配,本文使用一个简单的处理方法,将所得到的运动矢量按从小到大排序,对运动匹配块中运动矢量小的块先填充,对运动矢量大的块后填充,而对运动失配块最后进行处理。

图6(a)和图6(b)分别示意了对参考帧进行分割后的各个子块,以及对子块进行交叉叠块帧估计所得到的结果。图中,A、B、C、D分别代表对参考帧进行初始分割后所得到的4个子块,数字下标则表示对这些子块分别继续进行行和列分割所得的子块,数字从小到大表示分割进行的顺序。

其中,因缺少匹配覆盖而形成的“空洞”区域,如孤立点或单线,可通过邻域平均来进行填充;而对于面积较大的“空洞”区域,本文以一种只需要计算一次视差矢量的简单方法来进行填充:在右B帧图象中,首先设定一个面积较大的,包含了“空洞”区域和其四周填充区域的“窗口”,然后将这个“窗口”中的内容在左B帧图象中进行视差匹配估计,凡遇到位于“空洞”区域的点均不参与估计计算,最后得到视差矢量。根据视差矢量,将左B帧图象中的相应内容填充到右B帧图象中的“空洞”区域即可,从而克服了第1节方法的第4个缺陷。

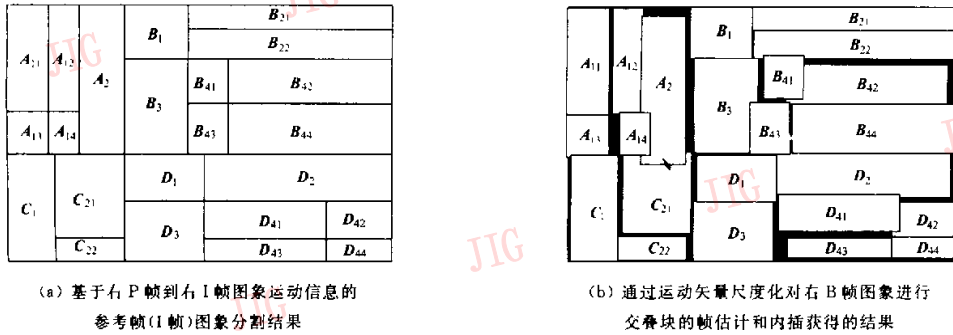


图 6 对参考帧分割及对交叠块帧估计示意图

### 3 实验结果

本文用 Booksale 和 Aqua 立体视频序列进行了仿真实验,并与文献[1]方法进行了比较.原始数据的每帧格式分别为  $360 \times 240$  和  $360 \times 288$ ,I 帧图象间的距离为 16 帧,I 到 P(或 P 到 P)帧图象之间有 3 个 B 帧图象.在多分辨率分割及运动估计过程中,使用了 3 个分辨率等级,并以 MAD 作为块匹配准则.

仿真实验结果见表 1,其中,左端 4 列是在解码端重建的可用以获得右 B 帧图象的参考帧图象的平均峰值信噪比 PSNR,右端二列为由这些参考帧图象经帧估计和内插所得右 B 帧图象的平均峰值

信噪比 PSNR,要说明的是这些 B 帧图象是每 3 个 B 帧图象中的中间一 B 帧图象,即性能最差的 B 帧图象.由表 1 可见,用本文方法所得到的信噪比,相对于文献[1]的方法提高了近 1.25dB.而在算法上两种方法的运动和视差搜索算法均采用全搜索方式,本文方法在时间上的开销约可节省 20%.

图 7、图 8 显示了用两种方法重建出来的中间 B 帧图象.由图 7、图 8 可知,用本文方法所得到的重建右 B 帧图象,尽管在运动区域也存在一些失真,但具有可接受的质量,而且由于分割点位于物体边缘处,所以在重建图象中,几乎没有恼人的“方块效应”.另外,经过 2018XL 立体显示器测试,重建的右 B 帧图象中,基本保留了与左 B 帧图象的视差(深度)信息.

表 1 两种算法的实验结果比较

序列名称	左 I 帧图象 PSNR(dB)	左 B 帧图象 PSNR (dB)	右 I 帧图象 PSNR(dB)	右 P 帧图象 PSNR(dB)	本文经帧估计和内插重建的右 B 帧图象 PSNR(dB)	文献[1]的方法得到的右 B 帧图象 PSNR(dB)
Aqua	36.78	33.57	32.30	32.03	28.89	27.62
Booksale	36.56	33.30	32.76	31.95	26.75	25.50

\* 因右 B 帧图象不需要编码传输,其 bpp 为 0.



(a) 原始的右 B 帧(第 80 帧)图象

(b) 用文献[1]方法得到的重建帧图象 (PSNR=25.61dB)

(c) 用本文方法得到的重建帧图象 (PSNR=27.70dB)

图 7 Booksale 序列



(a) 原始的右B帧(第22帧)图象

(b) 用文献[1]方法得到的重建帧图象  
(PSNR=25.39dB)(c) 用本文方法得到的重建帧图象  
(PSNR=28.64dB)

图8 Aqua序列

## 4 结论

对于立体视频序列的右图象流,本文的方法无需传输右B帧,可直接由解码得到的参考帧经帧估计和内插予以重建,并且由于采用了基于多分辨率四叉树分解的可变块分割方法和按照匹配块与失配块分别处理等原则,因而使重建右B帧图象的质量和搜索速度均比文献[1]的方法好.这表明本文方法是有效的.

由于重建的右B帧图象的失真主要还是发生在运动失配区,对因运动遮挡或新的背景区暴露而造成的估计失败问题尚需进一步研究,因此笔者正在研究MPEG-4的基于对象的运动检测以及通过建立一定的帧估计概率模型来进一步提高重建帧的速度和改善失真性能.

## 参考文献

- 1 Jiang Qin, Monson H Hayes III, Stereo Sequence Coding[J]. Visual Communications and Image Processing, 1998, 3309: 894~901.
- 2 Woontack Woo, Rate-distortion based dependent coding for stereo images and video: disparity estimation and dependent bit allocation[D]. PH.D Thesis, Dept. of Electrical Engineering, Faculty of the graduate school, University of Southern California, 1997:41~82.

- 3 Siegel M W, Sethuraman S, Mcveigh J S *et al.* Compression and Interpolation of 3D-Stereoscopic and Multiview Video[J]. Proc. of SPIE, 1997, 3012: 227~238.
- 4 Sethuraman S. Stereoscopic image sequence compression using multiresolution and quadtree decomposition based disparity-and motion-adaptive segmentation [D]. PH. D Thesis, Dept. of Electrical and computer Engineering, Carnegie Mellon University, Pittsburgh PA, 1996:41~67.
- 5 Sethuraman S, Siegel M W, Jordan A G. Segmentation based coding of stereoscopic image sequences[A]. In: proceedings of the IS&T/SPIE'S Symposium, on EI [C], San Jdoc, 1996:2668; 420~429.
- 6 Tekalp A M. 数字视频处理[M]. 北京:电子工业出版社, 1998: 196~212.



骆艳 1999年获清华大学学士学位,现为上海大学通信与信息系统专业博士生,研究方向为数字音视频处理和多媒体通信的数据压缩编码.

张兆扬 上海大学通信与信息工程学院教授,博士生导师.研究领域包括数字图象处理,视频信息压缩编码、处理、传输和多媒体通信等.发表论文90余篇,著书8部.