

一种运动自适应的帧速率上转换算法

杨越 高新波 李金秀

(西安电子科技大学电子工程学院 VIPS 实验室, 西安 710071)

摘要 在基于块的帧速率上转换算法中,块尺寸直接影响运动估计的准确性及插值帧的视觉效果。为此,提出了一种运动自适应的帧速率上转换算法。通过引入 STGS 图对视频内容进行预分析,根据运动特性自适应选取适合每一帧图像的块尺寸,进行重叠的块运动估计。并针对失真易产生在块边缘的特点,对块边缘的运动矢量进行插值细化处理,平滑运动矢量场,减轻人眼敏感的块效应。实验结果表明,该算法整体性能优于传统的固定块尺寸运动估计的帧速率上转换算法。

关键词 帧速率上转换 变尺寸块 STGS 图 运动矢量细化 运动估计 运动补偿

中图法分类号: TN919;TP391 文献标识码: A 文章编号: 1006-8961(2009)05-891-06

A Motion Adaptive Frame Rate Up-conversion Algorithm

YANG Yue, GAO Xin-bo, LI Jin-xiu

(Video and Image Processing Laboratory School of Electronic Engineering, Xidian University, Xi'an 710071)

Abstract In the block-based FRUC algorithm, block size directly affects the accuracy of motion estimation and the visual quality of interpolated frame. To this end, a new motion adaptive FRUC algorithm is proposed. Firstly, STGS image which analyzes the content of the frame is introduced to select a suitable block size adaptively for each frame according to the motion feature. Then motion vectors are obtained by using OBME (overlapped block motion estimate). Furthermore, the MVIR (motion vectors interpolated refinement) along the block boundaries is performed to reduce the block artifacts, which is capable of smoothing the motion vectors field and reducing block effect. Experimental results show better performance of the proposed algorithm compare with traditional block-based FRUC algorithm.

Keywords FRUC(frame rate up-conversion), changed size blocks, STGS(spatio-temporal gradient scale) image, motion vectors refinement, motion estimation, motion compensation

1 引言

随着数字电视及视频技术的蓬勃发展,大量视频格式标准并存。例如,电影节目播放速率为 24 帧/秒,而高清晰度电视(high definition television, HDTV)则要求 60 帧/秒以上的播放速率,因此,视频格式转换成为必不可少的环节。帧速率上转换(frame rate up-conversion, FRUC)能在两帧或多帧视频图像中产生处于中间运动状态的帧,

来实现不同的帧速率显示设备格式转换的要求。它广泛的应用在 HDTV、视频会议以及低比特率(low bit rates, LBR)编码中^[1]。目前,帧速率上转换算法主要分为两大类。一类为简单的方法,主要包括帧重复,时域线性平均^[2]。这类方法运算速度快,可以应用在实时性要求高的领域,但由于它们没有考虑物体的运动,所以在插帧中的运动区域会产生抽动、模糊等现象^[3-4]。为了解决这些问题,出现了第 2 类方法。这类方法考虑了运动信息,主要包括运动估计和运动补偿两个部分^[4-9],称为运动补偿帧

基金项目: Intel 中国研究中心合作项目;新世纪优秀人才支持计划项目(NCET-04-0948)

收稿日期: 2007-04-05; 改回日期: 2007-08-31

第一作者简介: 杨越(1983 ~),女,西安电子科技大学信号与信息处理专业硕士研究生。主要研究方向为图像、视频处理与通信。

E-mail: yangyue1983@126.com

速率上转换方法 (motion-compensated FRUC, MC-FRUC)。它可以有效地减少模糊和抽动等现象,为 FRUC 提供了很好的解决方法^[8]。但是运动估计 (motion estimation, ME) 的准确性直接影响着插值帧的效果,如果估计不够准确,在块的边缘容易产生明显的块失真^[5]。因此,很多算法都把注意力放在了获取物体真实运动矢量上^[10]。

在运动估计中,最常用的是基于块匹配 (block matching, BM) 的方法。它把图像划分为互不重叠的等分宏块,假设每个块内的运动是相同的,对每个块按照一定的标准进行匹配搜索,从而得到运动矢量。块匹配方法有位移跟踪能力强、运算复杂度低和易于硬件实现等优点,因此,得到广泛的应用。在 FRUC 算法中,已经提出了很多 BM 的方法^[2-9] 来获取运动信息,但是这些方法中每一帧块的尺寸是硬性划分的,并没有考虑不同运动类型的影响。同时每个块仅用一个运动矢量来描述运动信息,当块内包含多种运动时将得到不正确的运动矢量;又因块处理本身的缺陷,运动矢量连续性较差,导致在插值帧中块的边缘产生明显的块效应。

本文考虑了以上不足,给出了一种运动自适应的帧速率上转换算法 (motion adaptive FRUC, MA

FRUC),实现块尺寸的自适应划分和运动矢量的细化。实验结果表明,本文提出的方法得到的插值帧整体性能比固定尺寸块的方法得到了提高。

2 运动自适应的帧速率上转换方法

现有的可变块尺寸算法主要是在帧内选取块大小,对每一帧先划分为统一尺寸,然后进行运动估计,通过准则决定是否进行合并(分裂),最后重新计算运动矢量^[9],并没有考虑在帧间选择合适尺寸的块同样有着重要的作用。为此,本文提出了一种帧间块大小自适应选取的策略,通过计算 STGS (spatio-temporal gradient scale) 图,得到每一帧的运动复杂程度和静止百分比,根据提出的准则逐帧选取合适的块尺寸;同时针对块处理容易使物体边缘产生失真的缺点,采用了一种运动矢量细化的方法 (motion vectors interpolated refinement, MVIR)。利用现有的运动矢量,对块边缘矢量进行插值处理,这种将基于块的运动矢量变为基于像素的运动矢量的方法使物体边缘的突变现象明显减小。算法体系结构如图 1 所示。

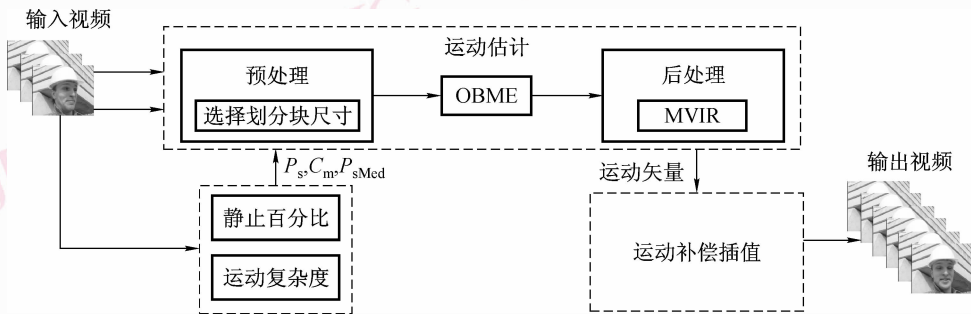


图 1 提出的运动自适应帧速率上转换的结构图

Fig. 1 The block diagram of the proposed motion adaptive FRUC algorithm

2.1 自适应块尺寸选取准则

运动估计前应该选择合适尺寸的匹配块。在块匹配算法中,因为小块易受噪声干扰,非常不稳定,选择大尺寸的块比小尺寸的块得到的运动矢量更加接近物体的真实运动^[5]。但是大块却不能很好描述物体的运动,如果块尺寸选择过大,有可能不满足块内像素运动一致性假设,忽略了小物体的运动,导致错误的运动估计。因此,块尺寸的选取至关重要。

帧间运动的剧烈程度对块大小选取有着一定的

影响,例如运动部分在每一帧中所占的百分比,各个块内运动一致性的好坏都可以作为块大小选取的标准。大量实验证明,帧间运动百分比,运动的复杂度可以较好地判定块大小的选择。基于 STGS 图的块尺寸的判决准则,通过对视频内容的预分析判断运动的复杂程度,进而自适应选取合适尺寸的块,改进了固定块的缺陷。

2.1.1 STGS 图的定义及物理意义

STGS 图包括水平 STGS 图 (h-STGS image) 和垂直 STGS 图 (v-STGS image)^[11]。分别定义为

$$S_h(x, y) = \frac{I_n(x, y)}{I_x(x, y)} = \frac{f_n(x, y) - f_{n-1}(x, y)}{f_n(x+1, y) - f_n(x, y)} \quad (1)$$

$$S_v(x, y) = \frac{I_n(x, y)}{I_y(x, y)} = \frac{f_n(x, y) - f_{n-1}(x, y)}{f_n(x, y+1) - f_n(x, y)} \quad (2)$$

其中, $I_n(x, y)$ 是 (x, y) 位置的时间梯度; $I_x(x, y)$ 和 $I_y(x, y)$ 分别是 (x, y) 位置的水平和垂直方向的空间梯度, f_{n-1} 、 f_n 是相邻两帧的像素值。

STGS 图的理论来自光流场的基本约束方程:

$$I_x u + I_y v + I_n = 0 \quad (3)$$

经过数学转化得到:

$$u = -\frac{I_n}{I_x} - \frac{I_y}{I_x} v \quad (4)$$

当帧间只有水平运动时,垂直分量 $v = 0$,此时水平分量为

$$u = -\frac{I_n}{I_x} \quad (5)$$

由式(5)可以看出,水平分量 u 与 h-STGS 的定义 $S_h(x, y)$ 只相差一个符号,同样 v-STGS 图也具有这样的关系。因为 STGS 图包含了两帧间的运动信息,所以它可以用来对视频内容进行预分析;同时 STGS 图的计算代价也很小^[11],因此,本文利用它来作为选取块尺寸的依据。

对于 STGS 图,根据式(1)、式(2)的符号,得到 STGS 的标记图,对其像素点的灰度 g 进行定义:

$$g(x, y) = \begin{cases} 0 & S(x, y) > 0 \\ 128 & S(x, y) = 0 \\ 255 & S(x, y) < 0 \end{cases} \quad (6)$$

这里, $S(x, y)$ 为水平或垂直的 STGS 图的像素值,灰度值 0、128、255 分别对应 STGS 图的黑、灰、白 3 色。STGS 图各点符号代表了该点的运动方向:正值代表该点向左或向上运动,负值代表该点向右或向下运动,零值代表该点静止。

2.1.2 基于 STGS 图的块尺寸选取准则

现实生活中的视频运动是非常复杂的,为了能够描述视频内容,把视频简单地分为 3 类:(1)具有静止背景的视频;(2)具有明显全局运动的视频;(3)具有复杂运动的视频。基于 STGS 图的块尺寸选取准则需要判定每一帧的运动属于上述 3 类运动中的哪一种,就必须分析这 3 类运动对应的 STGS 图各具有什么样的特征。表 1 给出了典型序列 grandma, garden, foreman 前 100 帧对应的 h-STGS 标记图和 v-STGS 标记图中黑色、白色、灰色像素点的平均数目。

表 1 STGS 标记图 3 种像素个数
Tab.1 The number of three kind of pixels in STGS image

序列	h-STGS			v-STGS		
	白色	黑色	灰色	白色	黑色	灰色
grandma	7 237	5 289	12 818	7 921	4 475	12 948
garden	65 894	19 779	15 703	56 869	27 120	17 387
foreman	11 597	7 432	6 315	10 415	8 310	6 619

大量实验证明,第 1 类视频对应的 STGS 图中静止像素点数目占的比例较大;第 2 类视频对应的表示运动的像素在某一方向上所占的比例要明显高于其他两种像素;第 3 类视频中表示运动的两种像素(黑色和白色)表示静止的灰色像素点的数目在两个方向上均相对平衡。在第 1 类视频中,大块能达到比较满意的效果,因此,只要运动不是太复杂,都选择了较大的块。第 2 类视频,选择中等尺寸的块较稳妥。在第 3 类视频中,各种块尺寸都有可能产生较好的效果,本文采用了多门限进行细分,达到选择最合适块尺寸的目的。

为了定量描述视频的内容,定义了 2 个变量,静止百分比 P_s^h 和运动复杂度 C_m^h :

$$P_s^h = \frac{N_{gray}^h}{M \times N} \quad (7)$$

$$C_m^h = \frac{1}{\frac{N_{max}^h}{N_{min}^h} - 1} \quad (8)$$

式中, $M \times N$ 为当前帧 $f_n(x, y)$ 的大小, N_{gray}^h 为 h-STGS 中灰色像素点个数, N_{max}^h 、 N_{min}^h 分别为 h-STGS 中黑色或白色像素点的个数的最大值与最小值。同理,对于 v-STGS 图也可以得到对应的 P_s^v 、 C_m^v 。

为了统计每一小段视频的运动特性,本文先在缓存内保存 25 帧图像,逐帧计算出 P_s 和 C_m ,这里 $P_s = \max(P_s^h, P_s^v)$, $C_m = \min(C_m^h, C_m^v)$ 。选取 25 帧中 P_s 的最小值 P_s^{min} 和最大值 P_s^{max} ,则均值为 $P_s^{med} = (P_s^{max} + P_s^{min})/2$ 。

根据静止百分比和运动复杂度先将缓存内视频每一帧分为前面提到的 3 类,如图 2 所示。静止百分比 $P_s > sh_1$,认为图像属于第 1 类运动; $P_s \leq sh_1$ 时,若运动复杂度 $C_m < mh_1$,认为属于第 2 类运动,否则判断为第 3 类运动。然后在每一类中进行不同尺寸的块的选择。

(1)判断为第 1 类运动:若 $C_m < mh_2$,选择块大

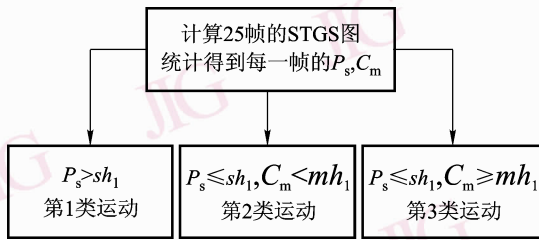


图 2 视频分类准则

Fig. 2 The video classification criterion

小 $size_1$; 否则, 选择块大小 $size_2$;

(2) 判断为第 2 类运动: 若 $P_s > sh_2$, 选择块大小 $size_2$; 否则, 选择块大小 $size_3$;

(3) 判断为第 3 类运动: 定义变量 $part = (P_s^{max} - P_s^{min})/4$, 门限 $th_1 = P_s^{med}$, $th_2 = th_1 - 2 \times part$ 。如果 $P_s > th_1$, 并且 $C_m < mh_2$, 选择块大小 $size_1$; 否则, 如果 $P_s > th_2$, 选择块大小 $size_2$; 否则, 选择块大小 $size_3$ 。

其中, sh_1, sh_2, mh_1, mh_2 均为经验参数; th_1, th_2 是根据视频特性的得到的参数; $size_1, size_2, size_3$ 是块的大小, 并且 $size_1 > size_2 > size_3$ 。

在运动估计之前通过上面的判别准则, 可以选择出适合每一帧大小的块, 根据视频内容, 达到最理想的划分效果。

2.2 运动估计

Taehyeun 等人提出了重叠的块运动估计 (OBME) 方法^[5]。该方法引入采样率来减小计算量, 并且使用 MWCI (modified weighted correlation index) 匹配标准, 得到的运动矢量更加准确。OBME 性能优于不重叠的 BME (block motion estimation), 因此, 本文采用 OBME 的方法得到运动矢量, 运动估计时采用变尺度块, 运动矢量插值细化及补偿时采用固定尺寸块。本文把块尺寸选取和 OBME 结合称为 AOBME (adaptive OBME)。

2.3 运动矢量的插值细化 (MVIR)

在基于块匹配的 FRUC 中, 存在“每个块内具有相同的运动”的假设。这个假设对含有非线性运动的块以及同时含有前景与背景的块是不成立的, 导致运动矢量错误, 插值帧内产生块失真。在块处理中, 失真通常产生在块的边缘^[6], 而且一部分失真是由于运动矢量的不连续性造成的, 称为块效应。H. 263 和 MPEG-4 已经提出了重叠的块运动补偿 (overlapped block motion compensation, OBMC) 算法来减小这种失真。这类方法主要是在运动补偿时对

当前帧和参考帧中的块进行了交迭、加窗、求和等处理。OBMC 是在运动补偿时对插值帧进行处理, 本文则提出一种运动矢量细化的方法, 在运动补偿前对运动矢量进行了平滑处理, 缓解了运动矢量的不连续性, 有效解决了块边缘的块效应问题。

如图 3 所示, 为 4 个互不重叠的块, 每个块均有自己唯一的运动矢量, 分别为 V_A, V_B, V_C, V_D 。在每个块的中心各划一个大小为原始块一半的小块, 如图虚线所示。图中的黑点称做大、小块的边界点。小块内的运动矢量保持不变, 为了减小块边缘的失真, 本文仅对小块以外的运动矢量进行处理。

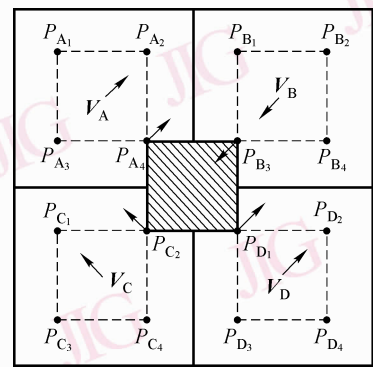


图 3 运动矢量插值

Fig. 3 Motion vectors interpolation

以边界点围成的阴影区域为例, 设 $P_{A_4}, P_{B_3}, P_{C_2}, P_{D_1}$ 的位置坐标分别为 $(x_1, y_1), (x_2, y_1), (x_1, y_2), (x_2, y_2)$, 在阴影块内任何位置 (x, y) 的运动矢量可以由边界点的运动矢量插值得到:

$$v(x, y) = abV_A + (1-a)bV_B + a(1-b)V_C + (1-a)(1-b)V_D \quad (9)$$

$$a = \frac{x_2 - x}{x_2 - x_1}, \quad b = \frac{y_2 - y}{y_2 - y_1} \quad (10)$$

边界点以外的区域都可以通过边界点的运动矢量按照式(9)、式(10)得到。块边缘的每个像素点都获得自己唯一的运动矢量, 平滑了运动矢量, 减轻了一个块仅用一个运动矢量表示带来的块效应。

2.4 运动补偿插值

运动补偿插值是利用运动矢量在两帧图像中插值处于中间运动状态的新帧。在前面的运动估计中, 处于块中间的小块只有一个运动矢量, 处于块边缘的像素都获得了各自的运动矢量。假设运动是线性的, 插值帧的运动处于当前帧和前一帧之间, 根据运动矢量, 得到插值帧 $f_{n-\frac{1}{2}}$ 为

$$f_{n-\frac{1}{2}}(x,y) = \frac{f_{n-1}\left(x + \frac{v_x}{2}, y + \frac{v_y}{2}\right) + f_n\left(x - \frac{v_x}{2}, y + \frac{v_y}{2}\right)}{2} \quad (11)$$

对于块, (x, y) 代表块坐标, 对于边缘的像素, (x, y) 代表该像素点的位置坐标。 $n-1, n-\frac{1}{2}, n$ 为时域坐标, (v_x, v_y) 为 (x, y) 处对应的块或者像素的运动矢量。 f_{n-1}, f_n 分别为参考帧与当前帧对应像素值。

3 实验与分析

本文设计了 5 组对比实验来验证提出的 MA FRUC 算法的有效性, 使用客观评价标准峰值信噪比 (PSNR) 来比较几种 FRUC 算法的性能。在实验中, 对原始的视频序列进行 2:1 的下采样, 在连续 3 帧中抽取中间帧作为要插值的帧, 其余两帧分别为参考帧和当前帧, 得到的插值帧与抽取出的帧进行主观和客观比较。

在所有的实验中, 搜索模式都使用全搜索 (full search, FS) 方式, 水平和垂直搜索范围均为 ± 16 个像素。在匹配准则的选取上, BME、OBME、AOBME 和 MA 均使用了 MWCI^[5] 匹配准则。块大小选取上, BME 使用 $size_3$, OBME 分别用了 $size_1, size_2$, AOBME 与 MA 块的大小利用准则来选取。实验中

取 $size_1 = 48 \times 48, size_2 = 32 \times 32, size_3 = 16 \times 16$ 。在 OBME, AOBME 和 MA 方法中, 块尺寸为 $size_1$ 时采样率为 3, $size_2$ 时采样率为 2, $size_3$ 时采样率为 1, 运动补偿插值时块的尺寸均为 16×16 。

使用了 6 组视频序列进行测试, 选用的视频主要分为 3 类, Grandma, Container 为背景静止, 前景只有微小运动的序列; Garden 是运动剧烈的全局运动序列。Carphone、Foreman、Suzie 为具有复杂运动的序列, 背景有时静止, 有时晃动, 人物面部运动比较大。对每一组视频使用 100 帧进行比较。参数 $sh_1 = 0.45, sh_2 = 0.1, mh_1 = 1.5, mh_2 = 3$, MWCI 中 $k = 0.02$ 。

从表 2 可以看出, AOBME 平均 PSNR 均高于 BME; 与 OBME 相比, 由于 AOBME 能够在 3 种块模式中选择适合视频内容的块尺寸, 看到 carphone、foreman、suzie 的平均 PSNR 值高于 OBME 中两种块尺寸下任何一种的平均 PSNR。Garden, grandma, container 的平均 PSNR 与 3 种块模式下最优值保持一致。还可以看出, 合适的块尺寸使得整体视频质量能有 0~1 dB 的改善, 但是固定的块大小却不能保证在每一种类型的视频中均取得最好的效果。AOBME 能选择在不同类型视频中选择大小合适的块, 其稳定性高于使用固定尺寸块。加入了 MVIR 后, MA 得到的插值帧质量比 AOBME 又有 0.02~0.6 dB 的改善。

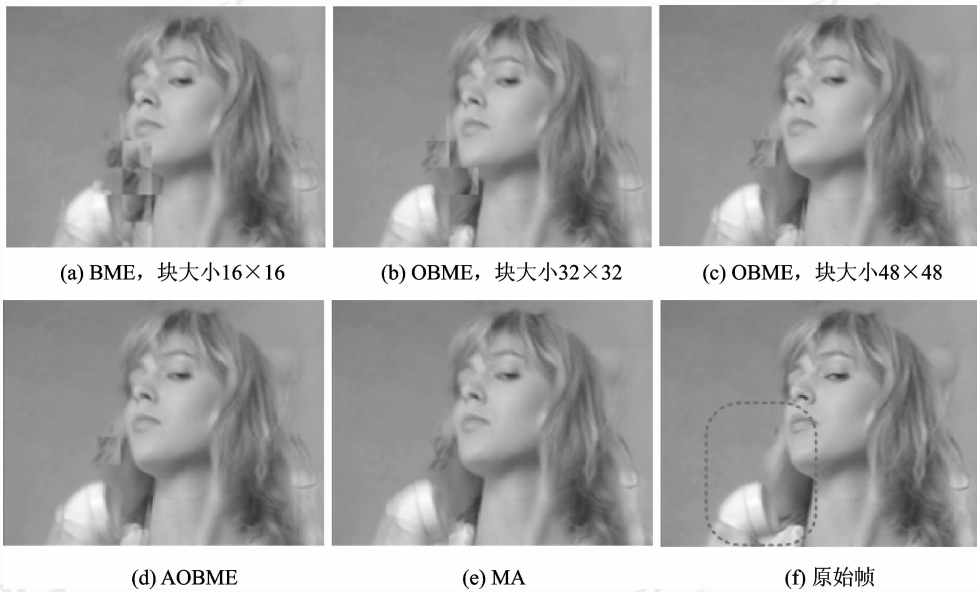
表 2 几种 FRUC 方法的插值性能比较

Tab. 2 Comparison of interpolation performance among several FRUC algorithm

块大小	BME	OBME		AOBME	MA	
	16 × 16	32 × 32	48 × 48	按准则选取	按准则选取	
PSNR (dB)	Carphone	32.921 4	33.386 4	33.298 4	33.526 0	33.583 8
	Foreman	32.015 5	32.227 8	32.042 5	32.234 0	32.896 8
	Suzie	36.021 3	36.238 7	36.100 0	36.279 8	36.860 7
	Garden	23.485 1	23.832 3	22.673 1	23.832 3	24.246 1
	Grandma	41.210 9	41.522 9	41.736 3	41.736 3	41.738 2
	Container	40.210 8	40.576 1	41.520 4	41.520 4	41.521 0
	平均 PSNR	34.310 8	34.630 7	34.561 8	34.854 8	35.141 1

采用了本文提出的 MA FRUC, 插值帧的主观质量也得到明显的改善, 图 4 给出了 suzie 序列第 54 帧的主观结果。注意观察在原始帧中用虚线标记区域的处理结果, 可以看出, 自适应块选取准则可以在

3 种块尺寸中选取主观质量较优的尺寸, suzie 选择了 48×48 的尺寸, 亦即选择了块效应最小的尺寸。同时经过 MVIR 处理, 脸旁边的块失真消失, 图像的质量也得到了改善。

图 4 测试序列 *suzie* 得到的插值帧Fig. 4 The interpolated frame of test sequences *suzie*

4 结 论

本文提出了一种新的自适应块尺寸选取准则,通过在帧间对视频内容进行分析对每一帧确定块的尺寸,避免了固定块尺寸统一划分的不足。同时,在得到运动矢量后对其进行了 MVIR,有效地减少了块边缘的失真,弥补了运动矢量的不连续性带来的不足。实验结果表明,本文提出的方法得到的插值帧在主观和客观质量都比传统的 FRUC 得到了提高,证明了所提出方法的有效性。

参考文献 (References)

- 1 Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. Draft ITU-T Recommendation and Final Draft International Standard of JointVideo Specification[S]. JVT-G050r1, 2003.
- 2 Pelagotti A, Haan D G. A new high quality algorithm for video format up-conversion[A]. In: Proceedings of the International Conference on Image Processing[C], Thessaloniki, Greece, 2001: 375-378.
- 3 Lagendijk R L, Sezan, M I. Motion compensated frame rate conversion of motion pictures[A]. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing[C], San Francisco, CA, USA, 1992: 453-456
- 4 Hilman K, Park H W, Kim Y. Using motion-compensated frame-rate conversion for the correction of 3 : 2 pull-down artifacts in video sequences

- [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, **10**(6): 869-877.
- 5 Taehyeun Ha, Lee S, Kim J. Motion compensated frame interpolation by new block-based motion estimation algorithm [J]. IEEE Transactions on Consumer Electronics, 2004, **50**(2): 752-759.
- 6 Lee S H, Kwon O, Park R H. Weighted-adaptive motion compensated frame rate up-conversion [J]. IEEE Transactions on Consumer Electronics, 2003, **49**(3): 485-495.
- 7 Bierling M, Thoma R. Motion compensating field interpolation using a hierarchically structured displacement estimation [J]. Signal Processing, 1986, **11**(4): 387-404.
- 8 Han S C, Woods J W. Frame-rate up-conversion using transmitted motion and segmentation fields for very low bit-rate video coding[A]. In: Proceedings of IEEE International Conference on Image Processing[C], Washington, DC, USA, 1997: 747-750.
- 9 Fujiwara S, Taguchi A. Motion-compensated frame rate up-conversion based on block matching algorithm with multi-size blocks [A]. In: Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems [C], Hong Kong, China, 2005: 353-356.
- 10 Haan D G, Biezen P W A C, Huijgen H, *et al.* True-motion estimation with 3-D recursive search block matching[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1993, **3**(5): 368-379.
- 11 Qi W, Zhang H J, Zhong Y Z. A pre-analysis method for robust global motion estimation [A]. In: Proceedings of IEEE International Conference on Image Processing [C], Kobe, Japan, 1999, **2**: 625-628.