

中图法分类号: TN919.81 文献标志码: A 文章编号: 1006-8961(2011)10-1763-09

论文索引信息: 杨春玲, 张兴绍, 莫望华. 像素级运动矢量场边信息生成算法 [J]. 中国图象图形学报, 2011, 16(10): 1763-1771

像素级运动矢量场边信息生成算法

杨春玲, 张兴绍, 莫望华

(华南理工大学电子与信息学院, 广州 510640)

摘要: 在分布式视频编码(DVC)中, 解码端生成的边信息的质量对整个系统的率失真(RD)性能有着重要影响。为了提高边信息的质量, 进而提升 DVC 系统的 RD 性能, 提出一种基于像素级运动矢量场的边信息生成算法(SGPMVF)。首先, 通过运动估计分别获得前后两关键帧的像素级运动矢量场; 然后, 为 Wyner-Ziv 帧的每个像素选择合适的运动矢量; 最后, 通过帧间内插得到高质量的边信息。算法中提出基于 SAD(绝对误差和)差值的搜索范围确定方法与像素级运动矢量选择方法。实验结果表明, 对于快速运动的序列, 该算法在不增加编码复杂度和码率的情况下使边信息的 PSNR 最大提高 1.45 dB, 并使 DVC 系统的 RD 性能提高 0.3~0.7 dB。

关键词: 分布式视频编码; 边信息生成; 像素级运动矢量场

Side information generation based on pixel level motion vector field

Yang Chunling, Zhang Xingshao, Mo Wanghua

(School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640 China)

Abstract: The quality of side information generated by a decoder has a significant influence on the Rate-Distortion (RD) performance of distributed video coding (DVC) codec. In order to improve the RD performance of a DVC by improving the quality of side information, a new side information generation algorithm based on pixel level motion vector field (SGPMVF) is proposed. First, the accurate pixel level motion vector fields (PMVF) of the forward and backward key frames are acquired respectively by motion estimation; then, the motion vectors of all pixels in the Wyner-Ziv frame are selected; finally, the high quality side information is obtained by interpolation. A new search range determination method based on the difference of SAD (sum of absolute difference) and a new pixel level MV (motion vector) selection method is proposed in the SGPMVF. Experimental results show that up to 1.45 dB gain for the side information and about 0.3~0.7 dB better RD performance can be achieved for fast motion sequences, while remaining the same rate and coding complexity.

Keywords: distributed video coding (DVC); side information generation; pixel level motion vector field (PMVF)

0 引言

DVC 作为一种新兴的视频编码框架, 具有编码简单、解码复杂、压缩效率高、鲁棒性好的特点, 非常

适用于编码终端的计算能力、存储能力及耗电量受限的应用场合, 例如无线视频监控、移动视频电话等。

DVC 的理论基础是 Slepian 与 Wolf 提出的分布式无损压缩理论^[1]及 Wyner 与 Ziv 在此基础上提出

收稿日期: 2010-04-23; 修回日期: 2010-12-15

基金项目: 国家自然科学基金项目(60972135); 中央高校基本科研业务费专项资金项目(2009ZM0231); 现代信息科学与网络技术重点实验室暨铁道部铁路信息科学与工程开放实验室开放基金项目(XDXX1009)。

第一作者简介: 杨春玲(1970—), 女, 教授。1999 年获南京理工大学电子工程系通信与信息工程专业博士学位, 主要研究方向为图像/视频压缩、图像质量评价。E-mail: eeclyang@scut.edu.cn。

的分布式有损压缩理论^[2]。在 DVC 的各种实现方案中,研究较多的是 Girod 等人提出的 Wyner-Ziv 视频编码(WZVC)系统^[3]。在 WZVC 系统中,解码端通过运动估计、运动补偿生成的边信息的质量直接影响到系统的 RD 性能。边信息与原始 Wyner-Ziv 帧越相近,解码端正确解码所需要的校验位越少,系统的编码效率越高^[3]。因此,生成高质量的边信息是提高 WZVC 系统性能的一个关键技术难点。

WZVC 系统中的边信息生成算法主要有两种:基于 Hash 码的运动补偿(MCH)算法和运动补偿外推/内插(MCE/MCI)算法。MCH 算法中,编码端将原始 Wyner-Ziv(WZ)帧 X 的某些特征信息变换成 Hash 码传送到解码端,辅助解码端进行运动估计生成边信息。由于利用了 X 的某些信息,该算法能生成较为准确的边信息,但是编码端要计算并传送 Hash 码,增加了编码端的复杂度和传输码率。MCE/MCI 算法能在没有任何原始 WZ 帧 X 信息的情况下在解码端生成边信息。MCE 算法的基本思想是,在 X 的前面两个已解码关键帧之间进行运动估计,然后基于线性运动假设外推出 X 中各图像块的运动,最后运动补偿生成边信息。由于仅利用了前面已解码关键帧,所以解码端可以进行顺序解码,但无法获得前后已解码关键帧之间的运动轨迹,生成的边信息质量不高。与 MCE 算法不同,MCI 算法^[4-6]利用 X 的前后两个已解码关键帧来估计 X 中各图像块的运动矢量,然后通过帧间内插生成边信息。MCI 算法充分利用时间相邻帧之间的运动信息,能在没有增加编码端复杂度和传输码率的条件下生成较准确的边信息。本文提出的基于像素级运动矢量场的边信息生成算法(SGPMVF)也属于 MCI 算法的范围。

在 MCI 算法中,有几个关键问题对算法的整体性能至关重要。

第 1 个问题是运动估计的分块尺寸选择。如果块的尺寸太小,容易使块匹配陷入局部最优;反之,则会导致运动矢量不够准确。为了解决这个问题,文献[7]提出一种分级运动估计(HME)算法。该算法首先使用尺寸较大的块进行运动估计以获得较粗糙的初始运动矢量;然后将块尺寸减半,利用仿射运动模型计算各子块的初始运动矢量,并自适应地计算各子块的搜索范围,再次执行运动估计对初始运动矢量进行优化,直到所获得的运动矢量场足够可靠、准确。文献[8]则利用 DVC 中运动矢量不用编

码传输的特点,提出一种差分运动估计(AME)算法,将 WZ 帧的块级运动矢量转换成稠密的像素级运动矢量,并通过最小化目标代价函数提高各像素运动矢量的精确度。

第 2 个问题是块匹配中搜索范围的确定。文献[7]提出一种基于邻近块运动矢量的自适应搜索范围计算方法,这种方法的有效性依赖于各块初始运动矢量的准确性,若某块的初始运动矢量是不准确的,则由它计算出来的相关块的搜索范围也是不准确的。为了更有效地利用初始运动矢量确定最佳搜索范围,本文提出一种基于 SAD 差值的搜索范围确定方法。

第 3 个问题是 WZ 帧中图像块(像素)运动矢量的选择。由于 MCI 算法在解码端进行运动估计时不能使用当前 WZ 帧的信息,所以要先估计出已解码关键帧的运动矢量场,然后从中为 WZ 帧的每一个图像块(或者每一个像素)选择合适的运动矢量。主要有以下 3 种方法:1)将已解码关键帧中相同位置块的运动矢量作为 WZ 帧中当前块的运动矢量^[6],这种方法最简单,但不够合理,也不精确;2)选择离 WZ 帧中当前块的块中心最近的运动矢量作为该块的运动矢量^[4];3)选择穿过 WZ 帧中当前块中心的运动矢量作为该块的运动矢量,对出现的多重覆盖区域采用平均法处理,对未覆盖区域再次进行运动估计^[9]。方法 3)以视频序列运动的线性平移为前提,而很多时候这种假设太过于简单而与真实的运动不相符,特别是对于运动比较复杂的序列。本文提出的像素级运动矢量选择方法,能从已解码关键帧的像素级运动矢量场中为 WZ 帧的每个像素(不是每个块)选择合适的运动矢量,从而得到能更加精确地代表原始 WZ 帧真实运动的像素级运动矢量场。

为了提高边信息的质量,本文提出一种基于像素级运动矢量场的边信息生成算法(SGPMVF)。该算法在获得前后已解码关键帧可靠的像素级运动矢量场的基础上,利用提出的像素级运动矢量选择方法为 WZ 帧的每一个像素选择合适的运动矢量,最后通过帧间内插与自适应加权平均得到高质量的边信息。

1 像素域 Wyner-Ziv 视频编码系统

像素域 Wyner-Ziv 视频编码系统^[3]如图 1 所

示。在编码端,视频帧被分成两种类型:关键帧 K 和 WZ 帧 X 。对于关键帧 K ,使用传统帧内编解码方法进行处理;对于 WZ 帧 X ,编码过程为:先对帧中的每个像素值进行 2^M 级均匀标量量化形成符号流 q ,将 q 进行码平面分离形成 M 个位平面,每个

位平面的二进制数据组成一个信息帧,然后用码率兼容删余 Turbo 码 (RCPT 码) 对每个信息帧独立进行编码。Turbo 码编码后只将校验位保存到缓存器中,根据解码器的反馈信息传送部分校验位至解码端。

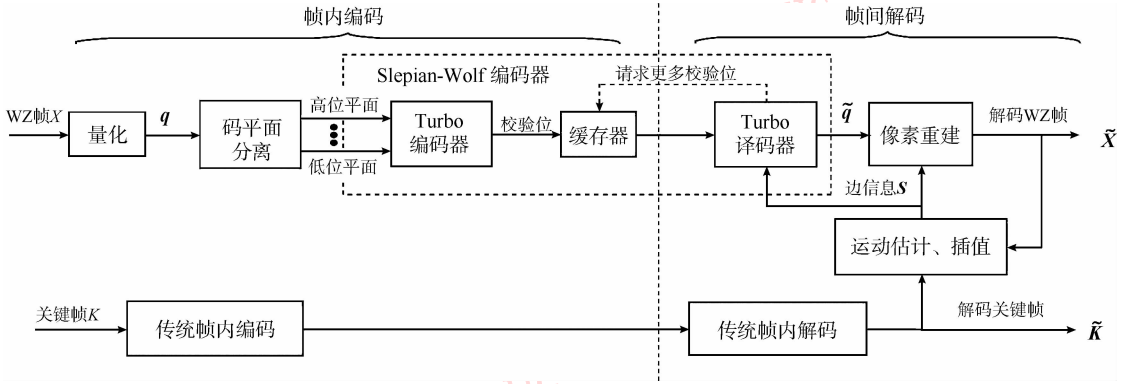


图 1 像素域 WZ 视频编码系统框图

Fig. 1 Architecture of pixel-domain WZ video coding

在解码端,利用已解码的时间相邻关键帧 \tilde{K} 进行运动估计、插值得到边信息 S 。Turbo 译码器利用编码器发送过来的部分校验位对边信息进行纠错译码。如果能正确译码,译码码流通过位平面重组得到符号流 \tilde{q} ; 否则向编码器请求更多的校验位重新译码,直到误码率小于指定的误码门限 P_e 。得到 \tilde{q} 后,解码器中的像素重建模块根据 \tilde{q} 和边信息 S 重构每个像素值,从而得到解码 WZ 帧 \tilde{X} 。

由解码过程可知,边信息 S 与原始 WZ 帧 X 越相似,正确解码所需要的校验位越少,系统的编码效率越高^[3]。所以,提出更加有效的边信息生成算法以提高边信息的质量在 WZVC 系统中是非常关键的。本文的创新工作体现在图 1 中的运动估计、插值部分。

2 像素级运动矢量场边信息生成算法

HME 算法^[7]与 AME 算法^[8]的一个共同特点是,从已解码关键帧 \tilde{K} 的块级运动矢量场中为 WZ 帧 X 的每一个图像块选择一个合适的初始运动矢量,然后通过一系列的后处理方法对这个初始运动矢量进行优化。由于物体的真实运动通常都不是简单的线性平移运动,并且一个图像块内可能包含多个具有不同运动的物体,所以获得的 \tilde{K} 的块级运动矢量场本身就不一定能准确地描述 \tilde{K} 的真实运动,

从中为 X 各块选择到的初始运动矢量跟实际运动矢量之间可能存在较大误差。由于 X 的数据是不知道的,HME 算法的优化过程以 X 中当前块的前向运动矢量与后向运动矢量的对称性为前提,但很多情况下这种假设太过于简单而与实际不相符,所以优化后的运动矢量也不一定能准确地表示 X 的真实运动,特别是对于运动复杂、剧烈的视频序列;AME 算法主要是通过增加一个小的修正值来对每个像素的初始运动矢量进行优化,当初始运动矢量与真实运动矢量之间的误差较大时优化效果不明显。

为了使估计出来的运动矢量场尽可能地接近真实的运动矢量场,从而提高边信息的质量,本文提出一种有效的基于像素级运动矢量场的边信息生成算法,结构框图如图 2 所示。与 HME、AME 等算法不同,该算法首先通过像素级运动矢量估计,分别获得当前 WZ 帧 X_{2i} 前后已解码关键帧 \tilde{K}_{2i-1} 、 \tilde{K}_{2i+1} 各自可靠的像素级运动矢量场 F_{2i-1} 、 F_{2i+1} ,然后利用提出的像素级运动矢量选择方法分别从 F_{2i-1} 、 F_{2i+1} 中为 X_{2i} 的每一个像素选择合适的运动矢量,得到能够更加精确地描述 X_{2i} 真实运动的后向和前向像素级运动矢量场 F_{2i}^B 、 F_{2i}^F 。再对 X_{2i} 的每个像素进行双向运动补偿,得到后向插值帧 P_B 与前向插值帧 P_F 。最后,为了充分利用 P_B 、 P_F 之间的互补性,也就是充分利用前后已解码关键帧之间的前向运动信

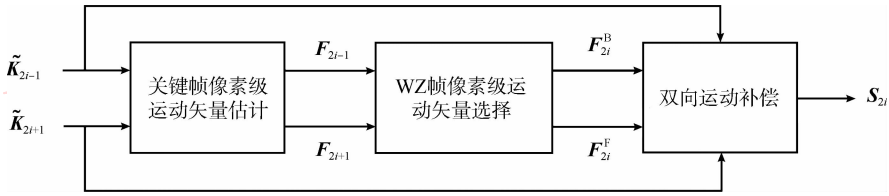


图 2 像素级运动矢量场边信息生成算法的结构框图

Fig. 2 Diagram of side information generation algorithm based on pixel-level motion vector field

息与后向运动信息,对 P_B 、 P_F 进行自适应加权平均,生成 X_{2i} 的高质量边信息 S_{2i} 。因此,所提算法包括已解码关键帧像素级运动矢量估计、WZ 帧像素级运动矢量选择、双向运动补偿和前后向插值帧的自适应加权平均 4 个内容,而前两个内容是本文算法的关键,具体描述如下:

2.1 已解码关键帧像素级运动矢量估计

已解码关键帧像素级运动矢量估计的目的是,

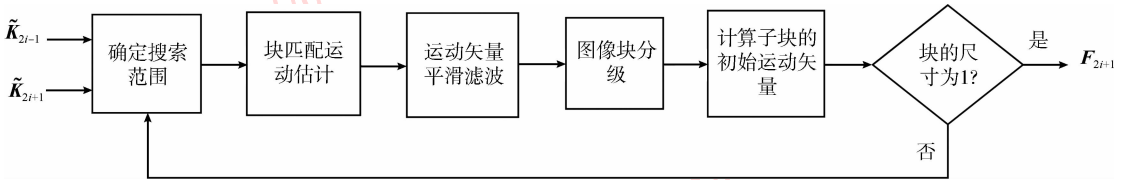


图 3 已解码关键帧 \tilde{K}_{2i+1} 像素级运动矢量的估计

Fig. 3 Pixel-level motion vector estimation of decoded key frame \tilde{K}_{2i+1}

2.1.1 基于 SAD 差值的搜索范围确定方法

在运动估计中,自适应地确定合适的搜索范围是十分必要的。

对于最大尺寸图像块(本文图像块的最大尺寸为 8×8),本文采用基于前帧运动矢量的搜索范围确定方法,利用视频序列运动的连续性确定当前已解码关键帧的搜索范围:若前面已解码关键帧的运动较剧烈,则将搜索范围设置为较大的范围,否则设置为较小的范围。

图像块分级后,图像块尺寸为 4×4 时,利用上一级较大尺寸块的运动矢量可获得当前级较小尺寸块的初始运动矢量。本文利用提出的基于 SAD 差值的搜索范围确定方法,根据获得的初始运动矢量场确定当前级各图像块的搜索范围。

基于 SAD 差值的搜索范围确定方法的基本思想是:对于初始运动矢量比较准确的图像块,只需在以初始运动矢量为中心的小范围内进行运动搜索;对于初始运动矢量不准确的块则应该在较大的范围内进行运动搜索,以纠正运动

获得当前 WZ 帧前后两个已解码关键帧每个像素的运动矢量。本文采用分级运动估计的思想,通过不断减小图像块的尺寸得到精细化的运动矢量,直至得到每个像素的运动矢量。图 3 所为关键帧 \tilde{K}_{2i+1} 像素级运动矢量估计的流程图,其中图像块的尺寸为 $L_k = \{2^k, k = K, K-1, \dots, 0\}$ 。提出的基于 SAD 差值的搜索范围确定方法及小尺寸块的初始运动矢量计算方法使得到的运动矢量更加精确。

矢量的错误。所以,方法的关键是如何判断一个运动矢量的正确与否。最简单的判断方法是根据该块与其在参考帧中的匹配块之间的 SAD 值来判定:如果 SAD 大于某一阈值,则认为该块的运动矢量是错误的。这种方法是不够准确的,由于各种反射和阴影效应,同一个物体点的成像亮度在物体运动前后可能发生变化,也就是说恒定亮度假设不总是正确的,所以尽管有时 SAD 比较大,但估计的运动矢量不一定是错误的。假设 E_{sad}^0 表示当前块与其在参考帧中的匹配块之间的匹配误差, $E_{sad}^i (i = 1, 2, \dots, 8)$ 表示当前块周围第 i 个邻块与其匹配块之间的匹配误差,本文提出的确定当前块搜索范围 r 的方法是

$$r = \begin{cases} r_1 & n > 4 \\ r_2 & n \leq 4 \end{cases} \quad (1)$$

式中, $r_1 > r_2$, n 表示 $E_{sad}^i (i = 1, 2, \dots, 8)$ 中与 E_{sad}^0 的差值大于阈值 T 的个数

$$n = \sum_{i=1}^8 a_i \quad (2)$$

$$a_i = \begin{cases} 1 & E_{\text{sad}}^0 - E_{\text{sad}}^i > T \\ 0 & E_{\text{sad}}^0 - E_{\text{sad}}^i \leq T \end{cases}, i = 1, 2, \dots, 8 \quad (3)$$

该方法的有效性基于以下两点:1)已解码关键帧中大部分 4×4 块的初始运动矢量都是比较准确的;2)如果两个相邻块的运动矢量都是正确的,那么即使外界条件发生变化,它们与各自在参考帧中的匹配块之间的 SAD 值也不会相差太大。所以,如果 E_{sad}^0 与 $E_{\text{sad}}^i (i = 1, 2, \dots, 8)$ 中多数的差值都比较大,则可判定当前块的运动矢量是不准确的;如果 E_{sad}^0 只与 $E_{\text{sad}}^i (i = 1, 2, \dots, 8)$ 中少数的差值比较大,则运动矢量错误的可能是邻域中的相关块而不是当前块。

当分级后块的尺寸为 2×2 时,各块的初始运动矢量已比较准确,只需对其进行微调即可,搜索范围设为很小的值。

2.1.2 计算子块的初始运动矢量

图像块分级后,一个大图像块分成 4 个小图像块,大块称为父块,小块称为子块。计算子块的初始运动矢量最简单的方法是直接采用父块的运动矢量。但是用这种方法得到的子块初始运动矢量是不准确的,因为当父块包含多个具有不同运动的物体(例如位于物体边界处的块)时,4 个子块的运动矢量之间可能有较大差异,而且可能与父块的运动矢量之间存在较大误差。文献[7]中提出的基于仿射运动模型的子块初始运动矢量计算方法,计算复杂度太高,且不能有效地解决精度问题。本文采用的方法是:从父块及与该子块相邻的 3 个大块的运动矢量中,选择出使该子块与其匹配块之间的 SAD 最小的运动矢量作为它的初始运动矢量。该方法利用实际运动场的局部相关性,能为子块选择到更加合适的初始运动矢量。

2.1.3 块匹配运动估计

为了提高运动矢量的准确性,文献[10]提出一种加权 SAD 准则。但是如果直接使用加权 SAD 准则来独立地最小化每一个图像块的匹配误差,得到的运动矢量场可能是比较混乱的,特别是当搜索范围比较大时。为了使得到的运动矢量场更加接近真实运动矢量场,本文将加权 SAD 准则^[10]与文献[5]中提出的带正则化项的匹配准则相结合,作为提出算法的运动估计准则。以 $\tilde{\mathbf{K}}_{2i+1}$ 运动矢量的估计为例,算法的运动估计准则如式(4)所示,在给定的搜索范围内寻找使代价函数 F_c 最小的运动矢量作为

$\tilde{\mathbf{K}}_{2i+1}$ 中当前图像块的运动矢量。

$$(v_x, v_y) = \arg \min F_c(d_x, d_y)$$

$$d_x \in [-M, M], d_y \in [-M, M] \quad (4)$$

$$F_c(d_x, d_y) = E_{\text{ksad}}(d_x, d_y) \times (1 + K \times \sqrt{d_x^2 + d_y^2}) \quad (5)$$

$$E_{\text{ksad}}(d_x, d_y) = \sum_{m=-t}^{L_k+t-1} \sum_{n=-t}^{L_k+t-1} \alpha(m, n) \cdot |X_{2i+1}(x+m, y+n) - X_{2i-1}(x+m+d_x, y+n+d_y)| \quad (6)$$

$$\alpha(m, n) = \begin{cases} 1 & 0 \leq m \leq L_k - 1 \\ & 0 \leq n \leq L_k - 1 \\ 0.5 & \text{其他} \end{cases} \quad (7)$$

$$t = \begin{cases} 2 & L_k = 8 \text{ 或 } L_k = 4 \\ 1 & L_k = 2 \end{cases} \quad (8)$$

式中, (v_x, v_y) 是获得的 $\tilde{\mathbf{K}}_{2i+1}$ 中当前块的运动矢量; M 是用 3.1.1 部分所述方法确定的搜索范围; K 是平滑约束参数,取为 0.05; (x, y) 、 L_k 为当前图像块的左上角像素坐标及其尺寸; t 是当前图像块外扩的像素数; $\alpha(m, n)$ 为文献[10]中给出的权值。

2.1.4 运动矢量平滑滤波

由于运动估计准则中没有考虑到相邻块对当前块的约束,所获得的运动场可能比较混乱,而真实的运动场从一个块到另一个块是平滑变化的(除了在物体边界处),所以不能保证得到的运动矢量都能够代表真实的运动,部分运动矢量可能是错误的。因此需要充分利用运动场的局部相关性对运动矢量作进一步的优化处理。本文采用文献[4]中的加权矢量中值滤波器对运动矢量进行平滑滤波,得到更加准确的运动矢量。设当前块 \mathbf{B}_0 的运动矢量及其八邻域块的运动矢量分别表示为 $\mathbf{mv}_i (i = 0, 1, \dots, 8)$, 如果 \mathbf{mv}^* 为这些运动矢量的中值,则有

$$\sum_{j=0}^8 \|\mathbf{mv}^* - \mathbf{mv}_j\| \leq \sum_{j=0}^8 \|\mathbf{mv}_i - \mathbf{mv}_j\|, i = 0, 1, 2, \dots, 8. \quad (9)$$

因此加权矢量中值滤波器的基本思想是,从 $\mathbf{mv}_i (i = 0, 1, \dots, 8)$ 中选择一个与其他 8 个运动矢量的加权距离和最小的运动矢量作为当前块 \mathbf{B}_0 的运动矢量。滤波器输出的运动矢量 \mathbf{mv}^* 满足

$$\mathbf{mv}^* = \arg \min_{\mathbf{mv}_i} \left\{ \sum_{j=0}^8 w_j \cdot \|\mathbf{mv}_i - \mathbf{mv}_j\| \right\} \quad (9)$$

$$i = 0, 1, 2, \dots, 8$$

$$w_j = \frac{E_{\text{sad}}(\mathbf{mv}_0, \mathbf{B}_0)}{E_{\text{sad}}(\mathbf{mv}_j, \mathbf{B}_0)} \quad (10)$$

式中, w_j 为运动矢量 mv_j 的权重; mv_0 为滤波前当前块 B_0 的运动矢量; $E_{\text{sad}}(mv_0, B_0)$ 为块 B_0 与由 mv_0 指向的匹配块之间的 SAD 值。显然, 候选运动矢量所决定的相应块之间的 SAD 值越大, 其权重越小, 对滤波器输出的影响越小, 反之亦然。

2.2 WZ 帧像素级运动矢量选择

WZ 帧像素级运动矢量选择的目的是, 分别从前后已解码关键帧的像素级运动矢量场 F_{2i-1} 、 F_{2i+1} 中为 WZ 帧 X_{2i} 的每一个像素选择准确的运动矢量。通常, 运动矢量选择方法都是基于线性平移运动假设, 然而真实的运动通常都不是简单的线性平移运动, 还有旋转、缩放等非线性平移运动 (例如人脸部的表情变化等), 这导致同一物体点在 3 个时间相邻帧中并不一定处于同一条直线上。本文提出的像素级运动矢量选择方法充分考虑到真实运动的复杂性, 选择出来的运动矢量能够更加准确地描述原始 WZ 帧的真实运动。

为了得到更高质量的边信息, 从 F_{2i+1} 中选择距离像素 p 最近的 2 个运动矢量作为当前像素 p 的有效运动矢量; 如果有多个运动矢量与像素 p 的距离相等且最小, 选择使前后已解码关键帧中对应的 2 个 3×3 图像块的加权 SAD 最小的 2 个运动矢量为有效运动矢量, 加权 SAD 值由下式表示:

$$E_{\text{ksad}}(mv_x, mv_y) = \sum_{k=-1}^1 \sum_{l=-1}^1 \alpha(k, l) \cdot \left| X_{2i+1} \left(m+k-\frac{mv_x}{2}, n+l-\frac{mv_y}{2} \right) - X_{2i-1} \left(m+k+\frac{mv_x}{2}, n+l+\frac{mv_y}{2} \right) \right| \quad (11)$$

$$\alpha(k, l) = \begin{cases} 5 & k=0, l=0 \\ 1 & \text{其他} \end{cases} \quad (12)$$

式中, (m, n) 为 X_{2i} 中像素 p 的坐标; (mv_x, mv_y) 为候选运动矢量; $\alpha(k, l)$ 为权值, 当前像素的权值较大, 邻近像素的权值较小。

类似地, 从 F_{2i-1} 中也为像素 p 选出两个有效运动矢量。这样为 X_{2i} 的每一个像素选择了 4 个运动矢量: 2 个来自 F_{2i+1} , 称为前向运动矢量, 分别表示为 mv_1^F 和 mv_2^F ; 2 个来自 F_{2i-1} , 称为后向运动矢量, 表示为 mv_1^B 和 mv_2^B 。为 X_{2i} 的每个像素选择多个运动矢量, 运动补偿时就可以为每个像素提供多个参考值, 能在一定程度上弥补线性平移运动模型的不足, 从而获得更高质量的边信息。式 (11) 采用加权 SAD 而不直接使用相关像素之间的差值, 能使得到

的运动矢量场更加平滑、可靠。

2.3 双向运动补偿

利用得到的前向运动矢量 mv_1^F 和 mv_2^F 对 X_{2i} 的各个像素进行双向运动补偿得到前向插值帧 P_F , 利用 mv_1^B 和 mv_2^B 进行双向运动补偿得到后向插值帧 P_B , 由以下公式计算得到

$$P_F(p) = (X_{2i-1}(p + mv_1^F/2) + X_{2i-1}(p + mv_2^F/2) + X_{2i+1}(p - mv_1^F/2) + X_{2i+1}(p - mv_2^F/2))/4 \quad (13)$$

$$P_B(p) = (X_{2i-1}(p - mv_1^B/2) + X_{2i-1}(p - mv_2^B/2) + X_{2i+1}(p + mv_1^B/2) + X_{2i+1}(p + mv_2^B/2))/4 \quad (14)$$

由于运动估计过程中不能使用原始 WZ 帧 X_{2i} 的数据, 而且视频序列的真实运动是很复杂的, 所以得到的 X_{2i} 的像素级运动矢量场中不免会存在一些不准确甚至错误的运动矢量。结果是有些像素在两个插值帧中都有较好的估计, 而有些像素只在其中一帧中有较好的估计, 甚至有些像素在两帧中的估计都不是很好。也就是说 P_F 与 P_B 之间具有一定的互补性, 综合利用 P_F 与 P_B 可以得到质量更好的边信息帧。本文将 P_F 和 P_B 中相同位置像素的自适应加权平均值作为边信息帧中像素 p 的估值, 即

$$S_{2i}(p) = \frac{E_{\text{sad}}^B(p)}{E_{\text{sad}}^B(p) + E_{\text{sad}}^F(p)} P_F(p) + \frac{E_{\text{sad}}^F(p)}{E_{\text{sad}}^B(p) + E_{\text{sad}}^F(p)} P_B(p) \quad (15)$$

$$E_{\text{sad}}^F(p) = |X_{2i-1}(p + mv_1^F/2) - X_{2i+1}(p - mv_1^F/2) + X_{2i-1}(p + mv_2^F/2) - X_{2i+1}(p - mv_2^F/2)| \quad (16)$$

$$E_{\text{sad}}^B(p) = |X_{2i-1}(p - mv_1^B/2) - X_{2i+1}(p + mv_1^B/2) + X_{2i-1}(p - mv_2^B/2) - X_{2i+1}(p + mv_2^B/2)| \quad (17)$$

3 实验结果及分析

对提出的 SGPMVF 算法进行仿真实验, 给出实验结果及分析, 并与文献 [6] 中提出的 SE-B 算法、文献 [7] 中提出的 HME 算法的结果进行比较。

3.1 实验条件

对于运动较慢的视频序列, 大多数边信息生成算法^[4,6-7] 都能得到较精确的边信息, 而对于运动较剧烈的视频序列, 现有算法得到的边信息质量不太

理想,为了验证本文提出的基于像素级运动矢量场的边信息生成算法,给出了运动强度中等的 Foreman 序列与运动剧烈的 Soccer 序列的实验结果。实验中对两个序列 299 帧的亮度分量进行编解码,奇数帧为关键帧,偶数帧为 WZ 帧。两个测试序列均为 QCIF 格式 (176 × 144),帧率为 30 帧/s,这意味着关键帧与 WZ 帧的帧率都是 15 帧/s。客观质量采用峰值信噪比 (PSNR) 值进行衡量。

实验结果包括两部分,一部分为生成边信息的性能比较,另一部分是基于 Turbo 码的 DVC 压缩性能比较。在 DVC 编码中,采用国际标准 H.264/AVC 的开放参考软件 JM16.2^[11]对关键帧进行帧内编解码,用解码的关键帧进行帧内插生成边信息。通过对 WZ 帧量化级数 $2^M \in \{2,4,8,16\}$ 的调整,产生不同的输出码率。对于不同的 M 值,采用不同的量化参数 (QP)对关键帧进行编解码,关键帧 QP 选择的准则是使 WZ 帧与关键帧具有近似的解码质量。系统所采用的 Turbo 编码器码率为 1/2,生成矩阵为 $\left[1, \frac{1 + D + D^3 + D^4}{1 + D^3 + D^4}\right]$,解码端误码门限 $P_e = 10^{-3}$ 。

采用文献[12]中提出的灵活可变重构算法,由解码的量化索引和边信息重构出最终的 WZ 帧。

3.2 实验结果及分析

1) 边信息的实验结果及性能分析

表 1 给出了当关键帧取不同 QP 时 3 种算法生成的边信息的质量比较。从表中可以看到,对于运动强度中等的 Foreman 序列,当关键帧的 $QP = 25$ 时,SGPMVF 算法比 SE-B 算法提高约 2.93 dB,比 HME 算法提高约 0.53 dB;对于运动剧烈的 Soccer 序列,当关键帧的 $QP = 26$ 时,SGPMVF 算法比 SE-B 算法提高约 3.41 dB,比 HME 算法提高约 1.45 dB。图 4 显示了当关键帧的 $QP = 25$ 时 3 种算法生成的 Foreman 序列所有边信息帧的 PSNR,图 5 显示了当关键帧的 $QP = 26$ 时 3 种算法生成的 Soccer 序列所有边信息帧的 PSNR。从上述实验结果中可以看出,对于两个视频序列,SGPMVF 算法生成的边信息的质量均优于其他两种算法,尤其是对于运动剧烈的 Soccer 序列。

表 1 不同算法生成的边信息平均 PSNR 值比较

Tab. 1 Average PSNR Comparison of side information generated by different algorithms

Foreman 序列				Soccer 序列			
关键帧 QP	SE-B	HME	SGPMVF	关键帧 QP	SE-B	HME	SGPMVF
25	32.861 1	35.259 3	35.788 0	26	24.887 6	26.839 5	28.294 1
29	31.591 6	33.778 3	34.185 9	33	24.190 5	26.320 2	27.540 5
32	30.474 8	32.501 9	32.824 6	37	23.674 9	26.039 3	26.979 3
35	29.205 5	31.151 6	31.344 7	40	23.148 4	25.380 3	26.406 6

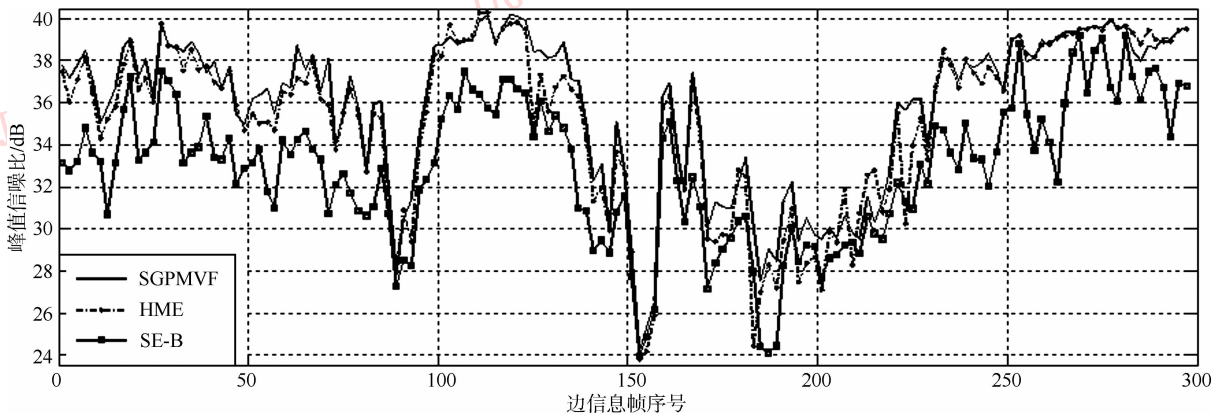


图 4 Foreman 序列边信息的质量

Fig. 4 Side information quality for the Foreman sequence

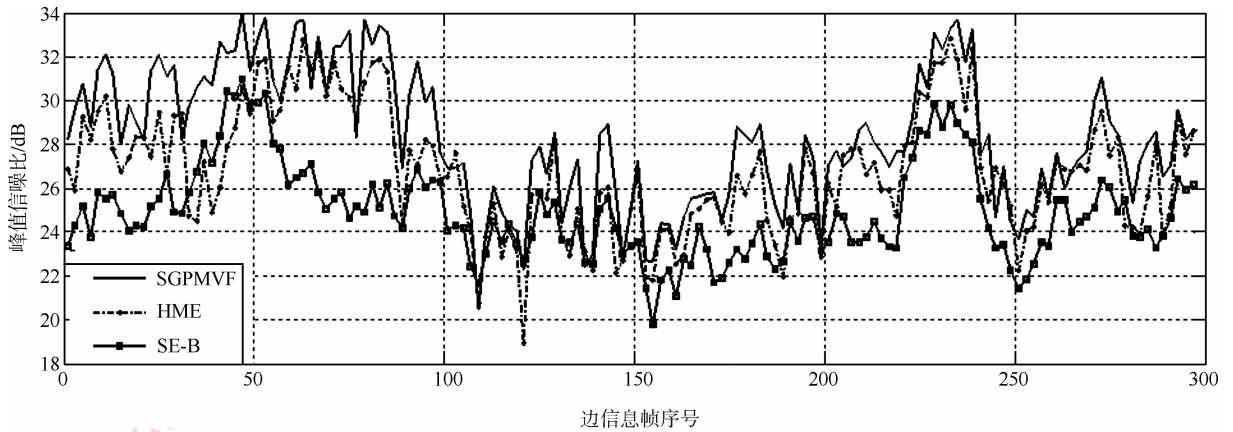


图 5 Soccer 序列边信息的质量

Fig. 5 Side information quality for the Soccer sequence

SGPMVF 算法对 Soccer 序列的改善效果比 Foreman 序列要明显,原因是 SGPMVF 算法中 WZ 帧的运动矢量场是从已解码关键帧可靠的像素级运动矢量场中获得的,所以当序列的运动很剧烈、复杂时仍然能够较准确地描述原始 WZ 帧的真实运动,从而获得更大的改善效果。

下面简要分析 SGPMVF 算法的复杂度。由于需要估计前后已解码关键帧每个像素的运动矢量,所以 SGPMVF 算法的计算复杂度要高于 SE-B、HME 算法。但是 SGPMVF 算法增加的仅仅是解码端的复杂度,而 DVC 系统更加关注的是解码端算法的性能而不是其计算复杂度。另外,SGPMVF 算法获得的已解码关键帧可靠的像素级运动矢量场有助于解码端利用已解码关键帧的有关信息对虚拟信道模型进行动态估计。

2) WZVC 系统的 RD 性能比较

为了评估本文提出的 SGPMVF 算法对 WZVC 系统 RD 性能的影响,本节将其应用到像素域 WZVC 中并与 H. 264/AVC Intra、文献[7]中提出的 HME 算法进行比较,率失真曲线如图 6、7 所示,表 2 列出了相应的具体实验数据。由于提出的 SGPMVF 算法提高边信息的质量,基于 SGPMVF 的 DVC 压缩性能也得到了改善。与 HME 算法相比,对于 Foreman 序列,系统的 RD 性能平均提高 0.2 dB;而对于 Soccer 序列,最多可提高 0.7 dB。可见,对于快速运动的序列,SGPMVF 算法能更加有效地提升 WZVC 系统的 RD 性能。同时也看到,像素域 WZVC 的性能与 H. 264/AVC Intra 相比有较大的差距,其原因是没有采用 DCT 变换等技术利用图像空

间相关性。

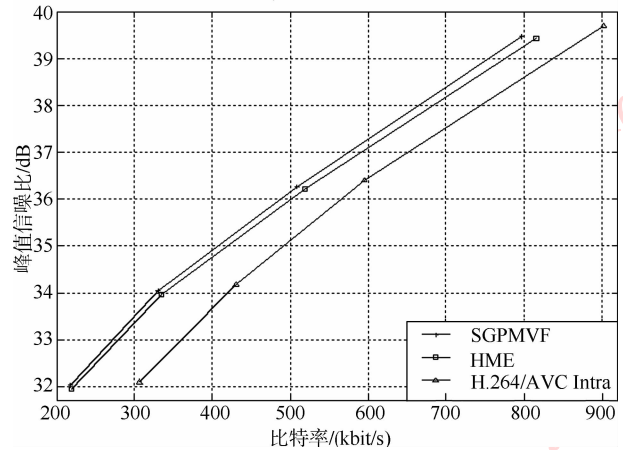


图 6 Foreman 序列 RD 性能对比

Fig. 6 RD performance for Foreman sequence

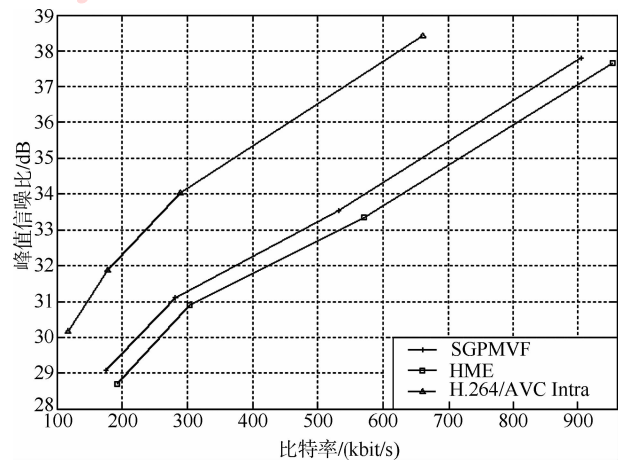


图 7 Soccer 序列 RD 性能对比

Fig. 7 RD performance for Soccer sequence

表 2 Foreman、Soccer 序列 RD 性能对比

Tab. 2 RD performance for Foreman and Soccer sequence

	K 帧 QP	HME		SGPMVF	
		Rate/ (kbit/s)	PSNR/ dB	Rate/ (kbit/s)	PSNR/ dB
Foreman	25	363.54	39.17	344.77	39.24
	29	220.20	36.02	210.46	36.12
	32	118.12	33.73	113.75	33.89
	35	64.70	31.74	63.46	31.90
Soccer	26	620.10	36.85	571.51	37.15
	33	425.75	32.61	386.12	33.01
	37	213.81	29.87	192.08	30.31
	40	132.83	27.20	115.47	27.93

4 结 论

生成准确的边信息是 DVC 系统中的一个关键技术。本文提出的基于像素级运动矢量场的边信息生成算法(SGPMVF),能够充分利用视频序列之间的运动信息,生成高质量的边信息,进而提升 DVC 系统的 RD 性能。实验结果表明,SGPMVF 算法对快速运动的视频序列压缩性能改善更加明显。将 SGPMVF 算法与迭代边信息生成算法相结合,研究两关键帧之间多个 WZ 帧的边信息生成算法是下一步的主要研究内容。

参考文献 (References)

[1] Slepian D, Wolf J. Noiseless coding of correlated information sources [J]. IEEE Transactions on Information Theory, 1973, 19(4): 471-480.

[2] Wyner A, Ziv J. The rate-distortion function for source coding with side information at the decoder [J]. IEEE Transactions on Information Theory, 1976, 22(1):1-10.

[3] Girod B, Aaron A, Rane S, et al. Distributed video coding [J]. Proceedings of the IEEE, 2005, 93(1):71-88.

[4] Ascenso J, Brites C, Pereira F. Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding[C]// Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services. Slovak: Slovak University of Technology, 2005.

[5] Ascenso J, Brites C, Pereira F. Content adaptive Wyner-Ziv video coding driven by motion activity [C]// Proceedings of ICIP2006. Piscataway, NJ: IEEE Press, 2006: 605-608.

[6] Li Z, Liu L, Delp E J. Rate-distortion analysis of motion side estimation in Wyner-Ziv video coding [J]. IEEE Transactions on Image Processing, 2007, 16(1): 98-113.

[7] Ascenso J, Pereira F. Hierarchical motion estimation for side information creation in Wyner-Ziv video coding[C]//Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication. New York, USA: ACM Press, 2008: 347-352.

[8] Cagnazzo M, Maugey T, Pesquet B. A differential motion estimation method for image interpolation in distributed video coding[C]// Proceedings of ICASSP2009. Washington, DC, USA: IEEE Computer Society Press, 2009: 1861-1864.

[9] Macchiavello B, Brandi F, Peixoto E, et al. Side-information generation for temporal and spatial scalable Wyner-Ziv codecs [J]. EURASIP Journal on Image and Video Processing, 2009, 2009:1-11.

[10] Gan Zongliang, Qi Lina, Zhu Xiuchang. Study of decoding algorithm for Wyner-Ziv video coding system [J]. Signal Processing, 2008, 24(4):609-613. [干宗良, 齐丽娜, 朱秀昌. Wyner-Ziv 视频系统中解码算法研究 [J]. 信息处理, 2008, 24(4):609-613.]

[11] Sühring K, Tourapis A. H264/AVC Reference Software J M16.0 [CP/OL]. [2010-04-14]. <http://iphome.hhi.de/suehring/tm1/download/>.

[12] Yang Chunling, Lin Yuguang, Po Laiman, et al. Variable reconstruction scheme for distributed video decoding [C]// Proceedings of Tencon 2008. Hyderabad: University of Hyderabad, 2008: 1-5.