

中图分类号: TN919.8 文献标志码: A 文章编号: 1006-8961(2011)01-0027-05

论文索引信息: 骆华燕, 朱秀昌. 时空域相结合的视频差错掩盖方法[J]. 中国图象图形学报, 2011, 16(1): 27-31

时空域相结合的视频差错掩盖方法

骆华燕, 朱秀昌

(南京邮电大学江苏省图像处理与图像通信重点实验室, 南京 210003)

摘要: 为了克服视频解码端时域差错掩盖技术不能准确估计丢失块运动矢量的缺点, 提出了一种将解码端运动估计和隐藏运动模型相结合的误码差错掩盖方法。首先, 利用丢失块周围正确接收像素估计丢失块运动矢量, 并计算估计的准确性。然后根据准确性, 通过隐藏运动模型充分利用丢失块的空域和时域相关性进行差错块的掩盖。仿真结果表明, 与现有的算法相比, 该方法能有效提高重建视频图像质量平均近 1 dB。

关键词: 视频编码; 差错掩盖; 运动估计; 虚拟相机; 隐藏运动模型

A temporal and spatial combined error concealment scheme for video communication

Luo Huayan, Zhu Xiuchang

(Jiangsu Province Key Laboratory on Image Processing & Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003 China)

Abstract: To overcome the shortcomings of temporal error concealment techniques in decoder that can not estimate motion vector of the loss block accurately, motion estimation in decoder and implicit motion model combined error concealment scheme is presented. First of all, we utilize the survived surrounding pixels of lost block for motion estimation, and determining the accuracy of estimates. Then we restore the missing pixels with implicit motion model making full use of the relevance of temporal and spatial domain according to the accuracy. The simulation results show that this method can improve the video image quality by 1dB than other methods.

Keywords: video encoding; error concealment; motion estimation; virtual camera; implicit motion model

0 引言

视频压缩编码标准 MPEG-x, H. 26x 主要使用了运动估计、运动补偿、离散余弦变换(DCT)和可变长编码(VLC)技术。这些技术在大量压缩视频数据的同时, 使得视频码流对信道误码非常敏感, 即使单个比特差错, 也可能造成恢复视频质量的急剧下降。差错掩盖是一种有效的抗误码手段, 利用视频序列时间和空间相关性及人眼视觉特性, 在解码端对误

码进行恢复, 提高视频质量。

空域差错掩盖技术是指丢失块和用来掩盖它们的数据来自同一帧, 但是它对于丢失块的细节恢复得不够。而时域差错掩盖利用视频的时间相关性, 用参考帧信息恢复差错块, 如视频序列中 P 帧和 B 帧, 可有效恢复图像细节。传统的时域差错掩盖尽可能估计出丢失块原始运动矢量(MV), 差别仅在于 MV 的估计算法不同。零运动矢量差错掩盖^[1] (也称时域替代, TR)是最早的时域差错掩盖技术, 只需将参考帧同一位置处图像块填补到当前帧丢失

收稿日期: 2009-07-07; 修回日期: 2009-09-03

基金项目: 国家自然科学基金项目(60672134)。

第一作者简介: 骆华燕(1986—), 女。南京邮电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处理、视频编解码与视频差错掩盖技术。E-mail: madfossil@126.com。

块的位置即可,无需任何计算量,但只适合静止图像序列,如新闻联播的背景画面。当运动变化较为剧烈时,效果很差。丢失块运动矢量为可用邻块运动矢量均值(AV)的恢复方法^[2]优于时域替代,但当丢失块与邻域块运动矢量相关性不强时,效果不好。Lam 提出的边界匹配算法(BMA)^[3],克服了丢失块的运动矢量与其周围块运动矢量相关性不强时掩盖效果不佳的缺点。但当丢失块的运动方向和周围块运动方向有不同步边界匹配算法产生严重的方块效应。

以上这些方法主要考虑的是丢失块运动矢量的恢复,其本身就有着一定的局限性。因为运动矢量的估计并不一定准确,即使能够准确估计出原始运动矢量,运动补偿后也极有可能会出方块效应。Li 提出了一种不需要估计 MV 的方法^[4],利用虚拟相机的概念,与隐藏运动模型相结合,得到了非常好的掩盖效果。在此基础上,提出了一种将解码端运动估计和隐藏运动模型相结合的方法。

1 解码端运动估计

如图 1 所示, t 时刻为当前帧图像,取出丢失块周围一圈正确接收到的像素。然后以丢失块位置为中心,在 $t-1$ 帧的一定区域内进行匹配,求出与丢失块临近的边界像素与前帧待匹配区域边界像素差的绝对差

$$MAD_1 = \frac{1}{N_{\text{pixel}}} \sum_{(i,j) \in E} |P_i(i,j) - P_{t-1}(i + V_{x1}, j + V_{y1})| \quad (1)$$

式中, E 为丢失块和待匹配块相邻的边界像素区域,如图 1 示, N_{pixel} 为边界像素总数。 $P_i(i,j)$ 和 $P_{t-1}(i + V_{x1}, j + V_{y1})$ 分别表示 t 帧丢失块和 $t-1$ 帧待匹配块边界像素值。最小的 MAD_1 被认为是最佳的匹配,它所对应的 (V_{x1}, V_{y1}) 即为估计出的丢失块的运动矢量。

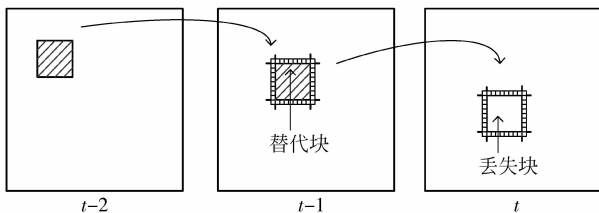


图 1 运动估计示意图

Fig. 1 Illustration of the motion estimation

传统的解码端运动估计(DMVE)^[5]算法就是将最佳匹配块直接填充到丢失块内,但是由于宏块周围的像素本身并不能保证处于一致的运动,因此以这些像素进行匹配得到的运动估计,很可能对于宏块中的某些像素是不准确的。本文得到 $t-1$ 帧的最佳匹配的替代块后,取出替代块并到 $t-2$ 帧中的一定搜索区域内进行匹配。同样,最小的 MAD_2 被认为是最佳的匹配,同时得到替代块的估计运动矢量 (V_{x2}, V_{y2}) 和最小的 MAD_2 。其中 M, N 为替代块在水平和垂直方向上的像素数。

$$MAD_2 = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |P_{t-1}(i,j) - P_{t-2}(i + V_{x2}, j + V_{y2})| \quad (2)$$

2 隐藏运动模型和时域形变

传统意义上的时域差错掩盖都是基于运动矢量的差错掩盖,都有可能因运动矢量估计不准确使得掩盖后出现块效应。Li 提出了一种不需要估计运动矢量的隐藏运动模型^[4],它首先估计出整帧图像的时域形变(temporal warping),然后以该形变为中心针对一帧图像构造出 9 个全局虚拟相机。这里所谓的虚拟相机就是指将帧与帧之间的时域形变看成是由于虚拟存在的相机的移动造成的。最后将 9 个虚拟相机掩盖结果进行融合得到最终结果。

2.1 AR 自回归模型

AR 自回归模型已大量用于 1 维 2 维信号,也有研究将其应用于 3 维信号并取得了较好的性能^[6]。根据图像局部平稳性质可知,一个像素可由一个局部的空-时域窗加权得到:

$$s(\mathbf{n}) = \sum_{i=1}^N a_i s(\mathbf{n}_i) \quad (3)$$

$\mathbf{n} = [x \ y \ t]^T \in [1, H] \times [1, W] \times [1, L]$ 表示视频序列中的某一点,它的空-时邻域用 N_n 表示, N_n 内像素点为 $\mathbf{n} = [x_i \ y_i \ t_i]$, $(i = 1, 2, \dots, N)$,如图 2 所示,即是以当前丢失块内其中某个像素点为中心的一个 $x \in [-1, 1], y \in [-1, 1], t \in [-1, 0]$ 邻域集合。因此,由式(3)可知,当前帧像素点可由其邻域 N_n 内像素点值加权得到。为了得到邻域内各像素点的加权值,构成一个局部训练窗 M_n ,如图 2,即以当前丢失块内其中某个像素点为中心的一个 $x \in [-2, 2], y \in [-2, 2], t \in [-1, 0]$ 邻域集合。

本文改进了 Li 的局部窗,使得当前帧的恢复只能用先前解码帧信息,目的是使其更加符合视频编解码系统的实际情况。

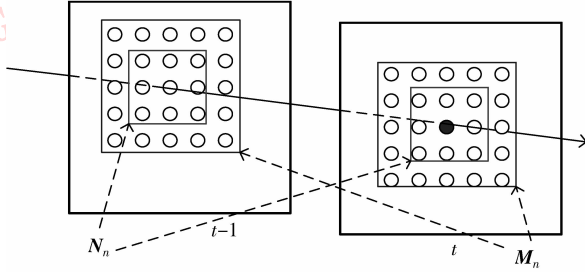


图 2 邻域 N_n 和 M_n 的定义

Fig. 2 Definition of N_n and M_n

与 2 维情况^[7]类似,利用最小化均方误差求 AR 模型系数

$$\mathbf{a} = (\mathbf{C}^T \mathbf{C})^{-1} (\mathbf{C} \mathbf{y}) \quad (4)$$

式中 \mathbf{C} 矩阵是检测集合 M_n 内有效像素点的 N_n 集合, \mathbf{y} 向量是 M_n 集合内有效像素点的集合。因此, \mathbf{C} 矩阵的构造会用到 $t-2$ 帧图像信息,也就是说当前帧丢失块的掩盖需要用到前两帧参考帧信息。此外,可用下式计算出每点的计算误差,用于后续的结果融合。

$$e = \|\mathbf{y} - \mathbf{C}\mathbf{a}\|^2 \quad (5)$$

2.2 针对每个丢失块的虚拟相机

Li 提出的 9 个虚拟相机是针对一帧图像而言的,它并没有充分考虑每个丢失块的运动特性。本文算法对每个丢失块进行考虑。对当前帧的某个丢失块,按照第 1 节方法计算其 MAD_1 , MAD_2 和 (V_{x1}, V_{y1}) , (V_{x2}, V_{y2}) 。设定阈值 Th ,用如下方法针对每个丢失块形成虚拟相机:

```

if ( $MAD_1 < Th$ ) 且 ( $MAD_2 < Th$ )
    则  $t-2$  帧虚拟相机为  $(V_{x2} + V_{x1}, V_{y2} + V_{y1})$ ;
     $t-1$  帧虚拟相机为  $(V_{x1}, V_{y1})$ ;
else
    则以  $(V_{x1}, V_{y1})$  为中心形成 9 个虚拟相机;
end

```

若 MAD_1 小于 Th 且 MAD_2 小于 Th ,则表明丢失块可以在前两帧中匹配得较好,宏块周围的像素极有可能处于一致的运动。这时,对当前丢失块而言,只需要一个虚拟相机即可,即将 $t-2$ 帧按照运动矢量 $(V_{x2} + V_{x1}, V_{y2} + V_{y1})$ 进行平移从而与当前帧保持一致。同理,将 $t-1$ 帧参考帧按照运动矢量 (V_{x1}, V_{y1}) 进行平移。将形变后的帧作为参考帧进行利

用,再用 2.1 节中的 AR 自回归模型掩盖当前丢失块。

若 MAD_1 大于 Th 或 MAD_2 大于 Th ,则表明在 $t-2$ 帧到当前帧 t 帧中,当前丢失宏块周围的像素本身并不能保证处于一致的运动。则以 (V_{x1}, V_{y1}) 为中心构造 9 个虚拟相机,这 9 个虚拟相机分别为 $(V_{x1} - 1, V_{y1} - 1)$, $(V_{x1} - 1, V_{y1})$, $(V_{x1} - 1, V_{y1} + 1)$, $(V_{x1}, V_{y1} - 1)$, (V_{x1}, V_{y1}) , $(V_{x1}, V_{y1} + 1)$, $(V_{x1} + 1, V_{y1} - 1)$, $(V_{x1} + 1, V_{y1})$, $(V_{x1} + 1, V_{y1} + 1)$,分别计算每个虚拟相机形变得到的掩盖结果。所谓虚拟相机的形变,就是 $t-2$ 和 $t-1$ 帧参考帧按照这些运动矢量进行平移从而与当前帧 t 帧保持一致。然后将平移后的帧重新作为参考帧,利用 2.1 节中的 AR 自回归模型掩盖当前丢失块。

针对每个丢失块构造虚拟相机进行时域形变的方法可以充分考虑每个丢失块的运动特性,而不是单纯地对整帧图像进行时域形变^[4],可在一定程度上提高掩盖质量。

2.3 虚拟相机掩盖结果的融合

得到了 9 个虚拟相机掩盖结果后,对其进行概率加权融合^[4]。由式(5)可计算出每点的计算误差,并用误差进行加权融合,权值为

$$\alpha^k = \frac{\exp\left(\frac{-e_k^2}{2\sigma_w^2}\right)}{\sum_{l=1}^K \exp\left(\frac{-e_l^2}{2\sigma_w^2}\right)} \quad (6)$$

式中, $K=9$,表示 9 个虚拟相机。 $\sigma_w^2 = 500$,是利用启发或探索法得到的一个常量值。

3 实验结果

为验证算法的效果,采用了 2 个不同的测试序列对本算法进行了仿真。CIF 格式的 mobile 和 foreman 序列。mobile 为物体的平移及来回移动,foreman 为物体移动和镜头一般性的旋转和平移。从序列的第 5 帧开始至第 38 帧,每隔 2 帧产生丢失块大小为 16×16 的误码,包括了独立丢失块和连续丢失块两种情况。2.2 节中的阈值通过实验探索法确定: $Th = 4$ 。若 Th 过小,则更多的丢失块将会形成 9 个虚拟相机,造成计算量增大。反之,则更多的丢失块只形成 1 个虚拟相机,造成虚拟相机不能正确刻画丢失块运动特性,使得掩盖效果不佳。最后利用时域经典算法及本文提出的算法进行差错掩

盖,并用客观量 PSNR 进行比较,这里只比较了 Y 亮度分量,并对这 12 帧 PSNR 结果进行平均,比较结果见表 1 和表 2。

表 1 独立丢失块几种方法 Y 分量平均 PSNR/dB 比较

Tab.1 The comparison of the average PSNR/dB of Y component for isolated block loss from different methods

序列	无遮盖	TR ^[1]	AV ^[2]	BMA ^[3]	Li ^[4]	本文
mobile	11.129 9	26.876 7	28.866 2	28.073 5	33.440 8	33.760 2
foreman	10.407 6	33.957 8	38.707 7	37.679 1	39.770 1	40.898 6

表 2 连续丢失块几种方法 Y 分量平均 PSNR/dB 比较

Tab.2 The comparison of the average PSNR/dB of Y component for consecutive block loss from different methods

序列	无遮盖	TR ^[1]	AV ^[2]	BMA ^[3]	Li ^[4]	本文
mobile	10.303 3	26.770 8	27.941 7	27.972 9	33.872 0	34.060 2
foreman	10.032 1	33.175 6	36.947 5	36.239 1	37.284 1	38.604 5

在 16 × 16 独立丢失块情况下,从表 1 可以看出,本文算法较传统算法 PSNR 有较大提高,相对于 Li 提出的算法而言,也有 0.3 ~ 1.1 dB 的增益。图 3 显示了独立丢失块情况下 mobile 序列第 29 帧



图 3 mobile 序列第 29 帧独立丢失块几种方法性能比较
Fig.3 The comparison of the result for isolated block loss from different methods of the 29th frame in mobile

用各方法进行遮盖的主观效果图,图 4 是对几种经典算法,从第 5 帧到第 38 帧各帧 PSNR 比较情况。可以看出,本文提出的算法具有较好的效果。

当丢失块大小为 16 × 16,且丢失块连续时,从表 2 可以看出,本文算法较 Li Xin 算法的 PSNR 值有 0.2 ~ 1.3 dB 的增益。图 5 显示了连续丢失块情

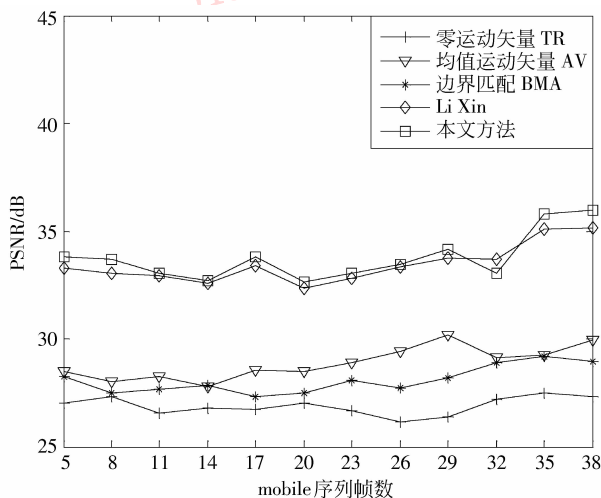


图 4 mobile 序列独立丢失块几种方法性能比较
Fig.4 The comparison of different error concealment results for isolated block loss in mobile sequence

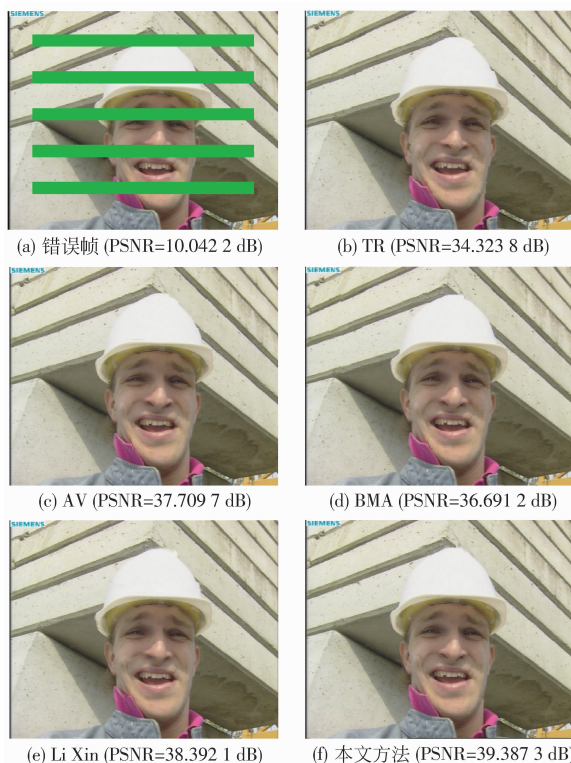


图 5 foreman 序列第 20 帧连续丢失块几种方法性能比较
Fig.5 The comparison of the result for consecutive block loss from different methods of the 20th frame in foreman

况下 foreman 序列第 20 帧用各方法进行掩盖的主观效果图,图 6 是连续丢失块情况下,各方法的 PSNR 比较,也可以看出本文算法的 PSNR 增益。

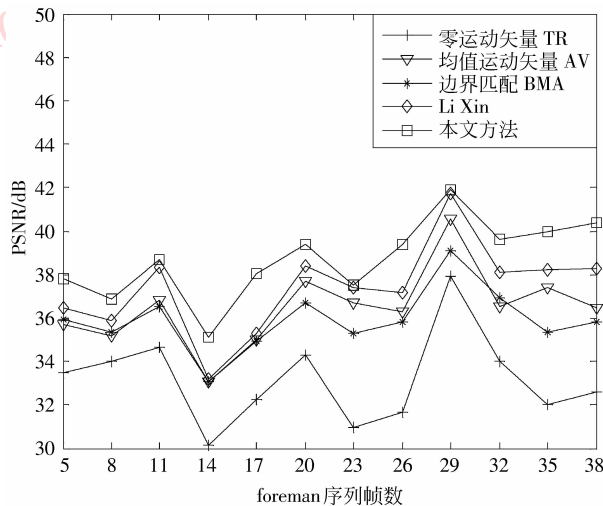


图 6 foreman 序列连续丢失块几种方法性能比较丢失

Fig. 6 The comparison of different error concealment results for consecutive block loss in foreman sequence

4 结 论

提出了一种将解码端运动估计和隐藏运动模型相结合的差错掩盖算法。针对每个丢失块进行局部虚拟相机的选择,改进了 Li 提出的整帧虚拟相机的方法,克服了时域差错掩盖技术不能准确估计丢失块运动矢量的缺点,回避了传统时域差错掩盖算法

单靠估计运动矢量的这一过程。从实验结果可以看出,本文算法不仅 PSNR 值提高平均近 1 dB,而且在主观质量上也取得了较好的效果。

参考文献 (References)

- [1] Aradhana Narula, Jae S Lim. Error concealment techniques for an all-digital high-definition television system [C]//Visual Communications and Image Processing' 93. Cambridge, MA, USA; Proc. SPIE, 1993: 304-315.
- [2] Keck W. A method for robust decoding of erroneous MPEG-2 video bitstreams [J]. IEEE Transactions on Consumer Electronics, 1996, 42(3): 411-421.
- [3] Lam W M, Reibman A R, Liu B. Recovery of lost or erroneously received motion vectors [C]//1993 IEEE International Conference on Acoustics, Speech and Signal Processing. Minneapolis, MN, USA; ICASSP, 1993: 417-420.
- [4] Li Xin. Video processing via implicit and mixture motion models [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(8): 953-963.
- [5] Zhang Jian, Arnold John F, Frater Michael R. A cell-loss concealment technique for MPEG-2 coded video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(4): 659-665.
- [6] Szummer Martin, Picard Rosalind W. Temporal texture modeling [C]//IEEE International Conference on Image Processing. Lausanne; ICIP, 1996: 823-826.
- [7] Li Xin, Orchard Michael T. Novel sequential error-concealment techniques using orientation adaptive interpolation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(10): 857-864.