

视频 Internet 传输的率失真优化运动估计算法

杨 静¹, 方向忠²

(1. 上海海事大学信息工程学院, 上海 200135; 2. 上海交通大学图像通信与信息处理研究所, 上海 200240)

摘 要: 通过对端到端失真的成因进行分析, 研究运动矢量的选择对误码扩散大小的影响。针对传统运动估计算法只考虑信源端的情况, 提出一种基于端到端预测误差的计算方法并据此对运动矢量进行率失真优化选择。基于 H.264 编码平台的仿真试验表明, 采用所提出的率失真优化运动估计算法可以进一步提高在 Internet 误码条件下的低码率视频流鲁棒性及解码质量。

关键词: 率失真优化; 运动估计; 低码率; 互联网

Rate-distortion Optimized Motion Estimation Algorithm for Video Transmission over Internet

YANG Jing¹, FANG Xiang-zhong²

(1. College of Information Engineering, Shanghai Maritime University, Shanghai 200135;

2. Institute of Image Comm. & Info. Processing, Shanghai Jiaotong University, Shanghai 200240)

【Abstract】 By analyzing the end-to-end distortion, the influence of motion vector selection on error propagation is studied. Allowing for that traditional motion estimation algorithm only takes into account the prediction error within source encoding framework, which is no longer accurate in lossy environment, an end-to-end prediction error calculation method is proposed. This prediction error is used in the rate-distortion optimized selection of motion vectors. Simulations have been done based on H.264 standard. The results show that for low-bit-rate video transmission over Internet, the robustness of video stream and the decoded video quality are improved by applying the proposed algorithm.

【Key words】 rate-distortion optimization; motion estimation; low-bit-rate; Internet

1 概述

近年来, 采用率失真优化方法对编码参数(如宏块编码模式, 量化参数, 运动矢量等)进行优化选择已经成为提高视频编码效率的重要手段之一。另一方面, 各种不同的视频流应用环境对视频的压缩和传输也提出了更高的要求。尤其是在基于 Internet 的视频流传输应用情况下, 常常会由于丢包现象从而在解码端产生严重失真。因此, 在率失真优化过程中, 通常在无误码环境下所采用的信源失真不再能够准确地反映误码环境下的实际失真。仅仅基于信源失真的率失真优化有必要被扩展到包括误码信道在内的基于端到端失真的率失真优化。

由于在编码时插入帧内刷新的宏块是阻断误码扩散的有效方法之一, 因此在许多文献中, 基于端到端失真的率失真优化方法被用于宏块编码模式选择上, 从而能够结合图像内容和信道情况适当地选择帧内刷新宏块的位置和数量, 取得了较好的抗误码效果^[1-4]。然而, 大量的研究只是集中在对宏块编码模式进行率失真优化选择从而达到抗误码的目的上, 对运动估计这个环节却很少有所关注。实际上, 在误码环境下, 尤其在低码率条件下, 由于码率的限制, 一帧图像中能用来进行帧内编码的宏块数量毕竟是有限的, 大部分的宏块还是帧间编码。采用帧内刷新来阻断误码扩散固然十分重要, 如果能设法削弱误码在帧间宏块间的扩散速度, 那么视频流的抗误码能力将会得到进一步提高。

2 Internet 信道下的误码失真扩散分析

假设视频流在 Internet 丢包信道下传输, 信道丢包率在

编码端已知, 在低码率下, 由于一帧图像压缩后的码字远小于 Internet 信道的 MTU 上限, 因此采用一帧直接作为一个 RTP 包的简单打包策略。如解码器发现当前解码帧所属数据包丢失, 则直接采用上一个解码帧的对应数据进行误码掩盖。

设 X_n^i 为编码帧 n 上像素点 i 的原始亮度值, \hat{X}_n^i 为在编码端重建的亮度值, \tilde{X}_n^i 为此像素在经误码信道后的解码重建值, e_n^i 为预测误差。采用误差平方和(summed square difference)为失真测度, 则此像素点端到端的总失真如下:

$$D_n^i = (X_n^i - \tilde{X}_n^i)^2 \quad (1)$$

且在编码端有下列关系成立:

$$\hat{X}_n^i = \hat{X}_{n-1}^j + e_n^i \quad (2)$$

其中, X_{n-1}^j 代表此像素在第 l 个参考帧的运动匹配参考点。

而在解码端, 当图像数据包被正确接收, 也就是运动矢量和预测误差为已知时, 有

$$\tilde{X}_n^i = \tilde{X}_{n-1}^j + e_n^i \quad (3)$$

当图像数据包发生丢失时, 直接采用上一解码帧的对应像素进行误码掩盖, 则有

$$\tilde{X}_n^i = \tilde{X}_{n-1}^i \quad (4)$$

考虑到 Internet 信道误码的随机分布特点, 解码端重建

基金项目: 国家“863”计划基金资助项目(2007AA12Z152)

作者简介: 杨 静(1973—), 女, 讲师、博士, 主研方向: 数字图像压缩编码与传输; 方向忠, 教授、博士生导师

收稿日期: 2007-04-06 **E-mail:** yangjinghao@hotmail.com

值 \tilde{X}_n^i 也是一个随机变量。假定在 $(1-p)$ 概率下图像数据包可被正确接收，而在 p 概率下数据包丢失并且在解码端采用误码掩盖，并且忽略信源失真 $(X_n^i - \hat{X}_n^i)^2$ 与信道失真 $(\hat{X}_n^i - \tilde{X}_n^i)^2$ 之间的相关性^[3]，则像素 X_n^i 的端到端总失真的数学期望为

$$\begin{aligned} E\{D_n^i\} &= E\{(X_n^i - \hat{X}_n^i)^2\} = \\ &E\{(X_n^i - \hat{X}_n^i)^2\} + E\{(\hat{X}_n^i - \tilde{X}_n^i)^2\} = \\ &(X_n^i - \hat{X}_n^i)^2 + (1-p)E\{(\hat{X}_{n-1}^j + e_n^j - \tilde{X}_{n-1}^j + e_n^j)^2\} + \\ &pE\{(\hat{X}_n^i - \tilde{X}_{n-1}^i)^2\} = (X_n^i - \hat{X}_n^i)^2 + \\ &(1-p)E\{(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)^2\} + pE\{(\hat{X}_n^i - \tilde{X}_{n-1}^i)^2\} \end{aligned} \quad (5)$$

从上式可以看出，总失真由 3 部分组成：第 1 部分为信源端的量化失真，第 2 部分是由运动匹配参考像素 X_{n-1}^j 所传递下来的扩散失真，第 3 部分是由误码掩盖所引起的掩盖失真。其中第 2 部分、第 3 部分可综合视为信道失真。在信道失真中， p 的值相对很小，因此扩散失真又占了主要部分。注意到这里的扩散失真是由参考像素 X_{n-1}^j 所传递下来，其大小与 X_{n-1}^j 的信道失真有关。由此想到，在误码信道下，运动矢量的选择对当前像素的信道失真大小有着重要的影响。如何优化选择运动矢量，使其指向失真较小的参考像素点，对于减小帧间误码的扩散进而提高视频流的传输鲁棒性有着重要的影响。因此，误码信道下的运动估计，不能仅以最小运动匹配预测误差为原则，还要综合考虑参考宏块所带来的信道失真，从而尽力使端到端意义上的总失真为最小。

3 基于端到端预测误差的率失真优化运动估计

假设采用误差平方和为失真测度，传统的理想信道下，运动估计的预测误差计算为

$$SSD(mv) = \sum_{i \in block} (X_n^i - \hat{X}_{n-1}^{i+mv})^2 \quad (6)$$

其中， mv 表示候选的运动矢量。

考虑误码传输环境后，将上式推广为端到端的形式，即

$$SSD_{channel}(mv) = \sum_{i \in block} (X_n^i - \tilde{X}_{n-1}^{i+mv})^2 \quad (7)$$

如第 1 节中所述，解码端重建值 \tilde{X}_n^i 是一个随机变量。而 X_n^i 和 \hat{X}_{n-1}^{i+mv} 在编码的运动估计阶段为已知，因此，忽略信源与信道之间的关联性，式(7)所示像素级端到端预测误差的数学期望可简化为以下直观和方便计算的形式：

$$E\{(X_n^i - \tilde{X}_{n-1}^{i+mv})^2\} \cong (X_n^i - \hat{X}_{n-1}^{i+mv})^2 + E\{(\hat{X}_{n-1}^{i+mv} - \tilde{X}_{n-1}^{i+mv})^2\} \quad (8)$$

如果用 D 表示信道失真，即

$$D_n^i = (\hat{X}_n^i - \tilde{X}_n^i)^2 \quad (9)$$

可以将式(7)重写为

$$SSD_{channel}(mv) \cong SSD(mv) + \sum_{i \in block} E(D_{n-1}^{i+mv}) \quad (10)$$

从上式中可看出，在误码环境下的预测误差为理想信道下的预测误差与候选运动矢量指向宏块所带来的信道失真之和。考虑到计算复杂度，在率失真优化运动估计中，采用绝对误差和(summed absolute difference)代替误差平方和，于是：

$$SAD_{channel}(mv) \cong SAD(mv) + \sum_{i \in block} E(d_{n-1}^{i+mv}) \quad (11)$$

其中， $SAD_{channel}(mv) = \sum_{i \in block} |X_n^i - \tilde{X}_{n-1}^{i+mv}|$ ；

$$SAD(mv) = \sum_{i \in block} |X_n^i - \hat{X}_{n-1}^{i+mv}|；d_n^i = |\hat{X}_n^i - \tilde{X}_n^i|。$$

则在误码信道下的率失真优化运动估计即为：选择运动矢量使式(12)所示的代价函数值为最小。

$$J = SAD(mv) + \sum_{i \in block} E(d_{n-1}^{i+mv}) + \lambda_{motion} R(mv) \quad (12)$$

其中， $R(mv)$ 表示编码运动信息所消耗的比特数。

信道失真 d 可采用迭代的方法进行计算。对帧内编码的像素，由于在 $(1-p)$ 的正确接收概率下，解码端不需运动补偿，因此信道失真的数学期望估计为

$$\begin{aligned} E\{d_n^i\} &= pE\{|\hat{X}_n^i - \tilde{X}_{n-1}^i|\} = pE\{|\hat{X}_n^i - \hat{X}_{n-1}^i + \hat{X}_{n-1}^i - \tilde{X}_{n-1}^i|\} \cong \\ &pE\{|\hat{X}_n^i - \hat{X}_{n-1}^i|\} + pE\{d_{n-1}^i\} \end{aligned} \quad (13)$$

对帧间编码的像素，信道失真的数学期望估计为

$$\begin{aligned} E\{d_n^i\} &= (1-p)E\{|\hat{X}_{n-1}^j + e_n^j - \tilde{X}_{n-1}^j - e_n^j|\} + pE\{|\hat{X}_n^i - \tilde{X}_{n-1}^i|\} \cong \\ &(1-p)E\{d_{n-1}^j\} + pE\{|\hat{X}_n^i - \hat{X}_{n-1}^i|\} + pE\{d_{n-1}^i\} \end{aligned} \quad (14)$$

根据式(13)和式(14)，每个像素的信道失真可在帧编码结束时进行迭代更新。

4 试验结果与分析

对 2 种率失真优化选择编码参数的方案在误码环境下进行了仿真。仿真时使用的打包及误码掩盖策略同上文所述。

方案 1 为仅在选择宏块编码模式时采用端到端率失真策略(以下简称 Mode 方案)，见文献[4]。方案 2 则是在方案 1 的基础上，进一步在运动估计环节采用所提出的率失真优化选择运动矢量的算法(以下简称 Mode&ME 方案)。

对 Foreman, Carphone 和 Silent 等 QCIF 序列在 H.264 JM86 平台上进行了仿真。编码参数均设为：帧率为 10 fps，序列第一帧为 I 帧，其余为 P 帧，编码 75 帧。仿真使用了 40~60 个不同的错误图样来模拟 Internet 误码信道，错误图样来源于文献[5]所提供的错误模型文件。总的 PSNR 值为码流通过不同错误图样丢包后所得解码结果的平均值。

图 1~图 3 为在典型丢包率 0.1 时，各测试序列使用 2 种方案后所得到的编码效率对比。

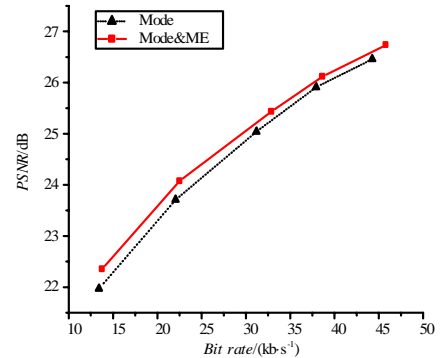


图 1 在 Foreman 中不同码率下平均 PSNR 值对比

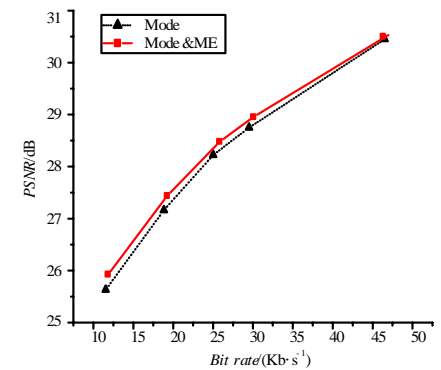


图 2 在 Silent 中不同码率下平均 PSNR 值对比

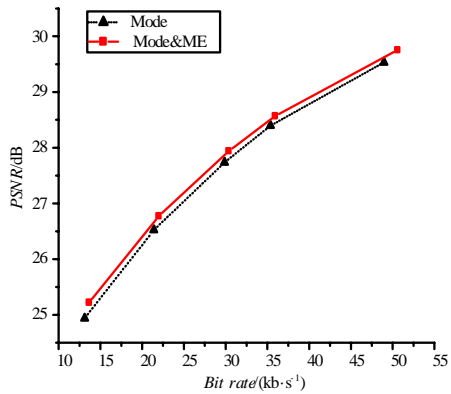
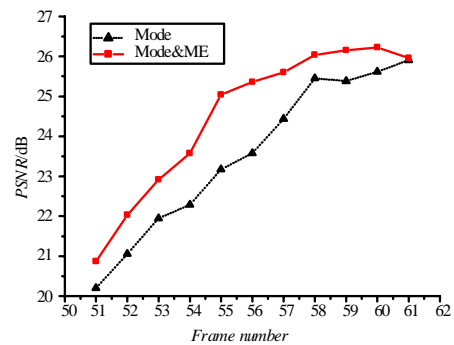


图3 在 Carphone 中不同码率下平均 PSNR 值对比

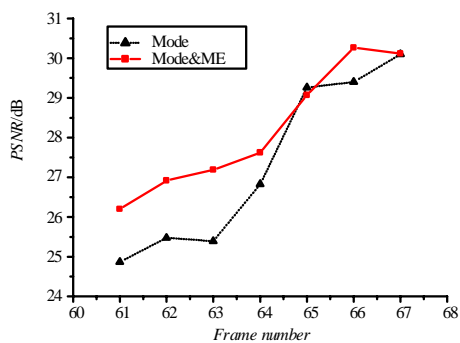
其中, 码率的调整是通过改变量化参数实现的。可看出 Mode&ME 方案与 Mode 方案相比取得了约 0.1 dB~0.4 dB 的 PSNR 增益, 且增益随着码率的增加而减小, 这是由于在低码率下, 没有足够的比特进行帧内编码, 率失真优化选择宏块编码模式的结果往往倾向于帧间模式, 可以说, 在低码率下, 采用帧内宏块进行误码抵抗的作用被削弱了, 这时, 如何选择合适的运动矢量使得误码扩散减小就变得更为重要。

总的看来, Mode&ME 方法所取得的 PSNR 增益并不十分明显。然而, 通过对不同的错误图样所取得的结果做进一步分析, 发现 2 种方案在很多丢包情况下所取得的结果是基本相同的, 但是当运动较为剧烈的图像帧发生丢失时, 2 种方案所取得的结果就产生了很大的区别。

图 4 绘出了当运动较为剧烈的图像帧发生丢失直到图像质量恢复正常时的一段 PSNR 曲线。Mode&ME 方案所取得的 PSNR 值明显比 Mode 方案高出很多, 最多可达 1.8 dB。但是由于仿真结果是多个错误图样结果的平均值, 因此总的平均 PSNR 增益却不太明显。



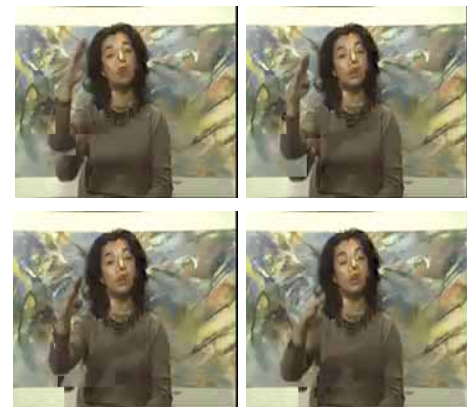
(a)Foreman



(b)Silent

图4 发生丢包后的 PSNR 值对比

图 5 为 Silent 序列发生丢包错误后连续几帧解码图像的直观视觉对比, 可见 Mode&ME 比 Mode 方案更快地恢复了图像质量。



(a)Mode



(b)Mode&ME

图5 发生丢包后的解码图像主观视觉对比

上述情况说明, 当图像包含较为剧烈的运动时, 图像帧丢失所引起的误码扩散也变得尤为严重, 但是由于码率的限制, 可以作为帧内编码的宏块数量是有限的, 而这时通过选择适当的运动向量, 使得运动向量指向失真较小的参考帧宏块, 从而进一步减少误码扩散, 所取得的效果是十分明显的。从图 1 中也可以看出, 运动复杂度较高的序列(如 Foreman)所取得的平均 PSNR 增益要比运动复杂度较低的序列(如 Silent)高。

5 结束语

本文研究了通过适当选择运动向量从而有效控制误码扩散的可能性。进而通过将基于端到端的预测误差引入到率失真优化运动估计环节中, 达到在误码环境下优化选择运动矢量的目的。在 H.264 JM86 校验模型平台上对所提出算法进行了仿真。结果表明, 采用所提出的率失真优化运动估计算法, 有效地提高了视频传输流的鲁棒性。尤其对低码率应用和运动复杂度较高的序列, 图像的解码质量在主观和客观上都有了较明显的改善。

参考文献

- [1] Zhang R, Regunathan S L, Rose K. Video Coding with Optimal Intra/Inter Mode Switching for Packet Loss Resilience[J]. IEEE Journal on Selected Areas in Communications, 2000, 18(6): 966-976.

(下转第 251 页)