

分布式视频编码虚拟依赖信道模型研究

房 胜^{1),2)} 李 哲¹⁾ 梁永全¹⁾ 钟玉琢²⁾

¹⁾(山东科技大学信息科学与工程学院 青岛 山东 266510)

²⁾(清华大学深圳研究生院 深圳 广东 518055)

摘 要 视频信息固有的非平稳特性,如冲突区域等,使时域预测技术变得非常复杂.在分布式视频编码(DVC)中,由于解码端不能获取当前编码帧的信息,精确地对时域相关噪声进行建模变得更为困难.文中以虚拟依赖信道模型为切入点,对如何降低时域相关噪声进行了研究.首先对DVC虚拟依赖信道进行了建模,并对影响边信息的主要因素进行了分析,分析结果表明在变换域中不同的频率子带对时域相关噪声的敏感度不同.在此基础上提出了一种新的基于小波变换域的虚拟依赖信道模型 VCMDWT,基于分类编码的思想对较为平稳的LL子带进行Wyner-Ziv编码,对非平稳的高频子带进行SPIHT帧内编码.实验结果表明,与基于像素域的方法相比,所提出的VCMDWT模型能够得到更稳定的虚拟信道,提高DVC系统的率失真性能达到2.6dBs以上.

关键词 分布式视频编码;虚拟依赖信道;信道模型;时域相关噪声;小波变换

中图法分类号 TP391

DOI号: 10.3724/SP.J.1016.2009.01404

Research of the Virtual Dependency Channel in Distributed Video Coding

FANG Sheng^{1),2)} LI Zhe¹⁾ LIANG Yong-Quan¹⁾ ZHONG Yu-Zhuo²⁾

¹⁾(College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266510)

²⁾(Graduate School of Shenzhen, Tsinghua University, Shenzhen, Guangdong 518055)

Abstract The inherent non-stationary characteristic of video signal, such as occlusion phenomena, leads to a complicated motion-compensated prediction technology. In distributed video coding (DVC) because the decoder cannot have access to the current frame, modeling the temporal correlation noise becomes a difficult task. In order to reduce the performance degradation owing to the non-stationary characteristic of video signal, this paper focuses on the research of virtual dependency channel model. Based on a common DVC codec, this paper presents a generalized model of the virtual dependency channel, analysis of which shows that in transformation domain the different subbands have various degree of sensitivity to the temporal correlation noise. According to the analysis this paper proposes a novel VCMDWT model, in which the virtual dependency channel is modeled at the LL subband with property of being more stationary, at the same time the non-stationary high frequency subbands are encoded by SPIHT with intra mode. The simulation results show that the proposed model is especially adaptable to the frames with many occlusion areas, which would be unfit for the pixel domain DVC, and finally improves the rate distortion performance with a gain up to 2.6dBs.

Keywords distributed video coding; virtual dependency channel; channel model; temporal correlation noise; wavelet transform

收稿日期:2007-05-09;最终修改稿收到日期:2008-10-20. 本课题部分得到国家自然科学基金重大研究计划“多方计算的安全保障机制和可信性检测与评价的关键技术研究”(90718011)资助. 房 胜,男,1971年生,博士,副教授,主要研究方向为视频编解码技术、多媒体通信、信息家电等. E-mail: fangsheng@tsinghua.org.cn; fang.sheng@sz.tsinghua.edu.cn. 李 哲,男,1983年生,硕士研究生,主要研究方向为视频编解码和通信技术. 梁永全,男,1967年生,博士,教授,博士生导师,主要研究领域为多媒体数据库、人工智能等. 钟玉琢,男,1938年生,教授,博士生导师,主要研究领域为多媒体数据压缩、流媒体、数字化家电网络平台.

1 引言

随着无线网络和移动视频设备的迅速发展,一种新型的“上行(Up-Link)”视频应用需求正在出现,与传统的面向“下行(Down-Link)”应用的视频编码标准 MPEG-X 和 H.26x 等不同,它要求编码器简单,但允许解码器具有高复杂度^[1]. 目前分布式视频编码 DVC(Distributed Video Coding)技术被认为是“上行”视频应用各种解决方案中最有前景的. DVC 主要从 Slepian、Wolf 建立的分布式无损编码理论^[2]和 Wyner、Ziv 建立的分布式有损编码理论^[3]发展而来.

分布式信源编码 DSC(Distributed Source Coding)是指对两个或者多个相关的随机过程进行编码,每一个过程使用一个单独的编码器,如图 1 所示. Slepian-Wolf 定理表明这种编码场景存在如下关系^[2]:

$$\begin{aligned} R_X + R_Y &\geq H(X, Y), \\ R_X &\geq H(X|Y), \\ R_Y &\geq H(Y|X) \end{aligned} \quad (1)$$

式(1)令人惊异地表明了虽然 X 和 Y 是分别编码的,但是总码率 $R_X + R_Y$ 可以与联合信息熵 $H(X, Y)$ 相等,这与对 X, Y 进行联合编码的情况是相同的.

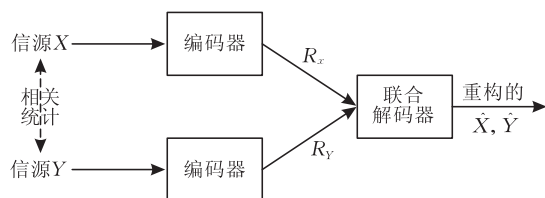


图 1 Slepian-Wolf 无损分布式编码示意图

Wyner-Ziv 编码器可以看作是由一个量化器后紧跟一个 Slepian-Wolf 编码器组成的,如图 2 所示. Wyner 和 Ziv 证明了 $R_{X|Y}^{WZ}(D) \geq R_{X|Y}(D)$, 其中 $R_{X|Y}^{WZ}(D)$ 表示在编码端不能获得边信息(side information) Y 情况下的 Wyner-Ziv 率失真函数, $R_{X|Y}(D)$ 表示编码端、解码端都可以获取边信息 Y 情况下的率失真函数^[3]. 该公式的成立为 DVC 技术奠定了理论基础,它表明了一个不同于传统视频编码技术的编码模式,即编码时每帧独立编码,而解码时利用边信息进行联合解码. 该技术对于移动视频应用来说具有巨大的吸引力,因为编码时只采用帧内编码方式,ME/MC 等高计算复杂度的模块则被转移到解码端进行,复杂度大为降低,而在理论上

又能获得与传统编码技术相同的编码效率.

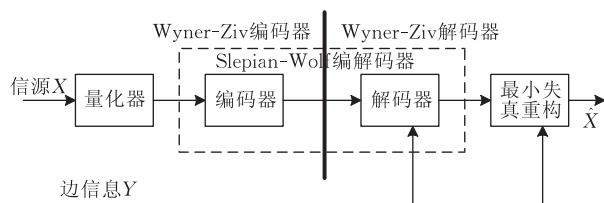


图 2 分布式视频编码结构

在图 2 所示的分布式视频编码模型中,边信息 Y 与当前编码信息 X 之间的依赖关系可以视作是一个虚拟依赖信道, Y 则被看作是 X 经由信道发送而产生错误的噪声污染版本. 解码过程就是利用信道编码的校验信息和噪声污染版本 Y 恢复出原始信号 X . 由于视频信息固有的非平稳特性和解码端不能获取当前帧信息等原因,构造的边信息 Y 的质量是动态变化的,导致 X 与 Y 之间的虚拟依赖信道呈现非平稳特性,随编码视频内容的变化而变化. DVC 系统的性能与信道编码的效率紧密相关,而无论采用哪种信道编码技术,其编码性能都依赖于虚拟依赖信道模型的准确度. 如果采用的虚拟依赖信道模型不能正确反应 X, Y 之间的统计相关性,那么编码的效率就会大幅下降. 目前国内外的研究文献中对 DVC 虚拟依赖信道的相关研究基本上是在像素域进行的,主要是研究如何通过 MEMC 生成更为精确的边信息^[4-7],动态调整虚拟依赖信道的模型参数^[8],或者通过建立不同区域的虚拟依赖信道模型提高编码效率^[1,9].

本文对 DVC 中的虚拟依赖信道模型进行了研究,目标是尽量降低虚拟依赖信道非平稳特性引起的 DVC 编码质量的下降. 根据建立的虚拟依赖信道模型,对边信息的构造方式、边信息的预测域等问题进行了理论分析,并提出了一种新的基于小波变换域的虚拟依赖信道模型 VCMDWT. 本文第 2 节对虚拟依赖信道进行分析,并建立一个统一的模型;第 3 节对该模型进行研究分析;第 4 节是实验结果;第 5 介绍相关工作;最后是本文的结束语.

2 DVC 系统中的虚拟依赖信道

图 3 中描绘的是目前 DVC 研究中比较通用的一种框架结构^[4-12],其中解码端生成的边信息 Y 和编码端的编码信息 X 构成了一个虚拟依赖信道. 根据公式(1),即 Slepian-Wolf 定理,DVC 系统编码的最低码率由 $H(U|Y)$ 唯一决定^[2]. 理论上能够无差

错重构原始信号的最低码率 R 为

$$\begin{aligned} R &\geq H(U|Y) = \sum_y P(y) H(U|y) \\ &= \sum_y P(y) \sum_u P(u|y) \log P(u|y) \quad (2) \end{aligned}$$

其中 $P(u|y)$ 可以由 $P(x|y)$ 计算得到. 式(2)表明了虚拟依赖信道模型对于 DVC 码率的决定性影响. 视频信息可以看作是一个 3D 随机场的实现过程, 它随着时间的变化而变化. 解码端在不知道当

前帧 X 的情况下, 重构出的边信息 Y 是否满足编码端假设的信道概率模型是不确定的. 在编码器不能获得确切的边信息 Y 的情况下, 它只能假设概率模型已知, 并按着已知的概率模型工作. 当生成的 Y 与 X 的相关性较差时, Y 与 X 之间的误码率可能会超过信道编码的校验能力, 此时 DVC 编码的效率就会急剧下降.

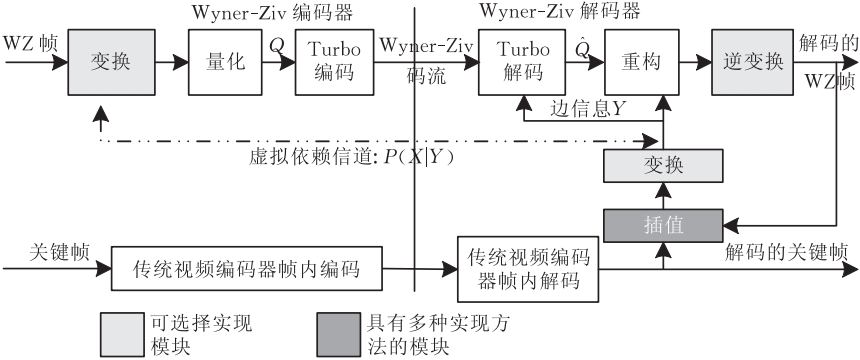


图 3 分布式视频编码结构及其虚拟依赖信道

边信息 Y 通过对已经解码的视频帧进行估计而生成. 一个通用的虚拟依赖信道模型如下式所示, $Y(i, j, t) = f(X_{\text{decoded}}(i, j, t+1), X_{\text{decoded}}(i, j, t-1), \dots,$

$$\begin{aligned} &X_{\text{decoded}}(i, j, t-N)) \\ &= X(i, j, t) + N(i, j, t) \quad (3) \end{aligned}$$

其中 i, j 表示系数空间坐标, t 表示时域坐标, f 表示使用前 N 个已经解码的视频帧和下一时刻解码的视频帧生成边信息的函数^①. 如果要对虚拟依赖信道进行准确的描述, 就必须能够准确地描述 $N(i, j, t)$. 在实际的通信信道中, 真实物理信道的模型一般是已知的. 在 DVC 中, 由于原始信号 X 和边信息 Y 的非平稳特性, 由 $N(i, j, t)$ 决定的虚拟依赖信道难以找到对其进行准确描述的分析模型.

3 虚拟依赖信道模型分析

根据式(3)可知, 由于时域相关噪声 $N(i, j, t)$ 的存在, 对 DVC 虚拟依赖信道在时域进行建模非常困难. 从图 3 可以看出, 影响虚拟依赖信道模型的主要因素有原始信息 X 和边信息 Y . 作为输入的原始信号 X 不能被改变^②, 而 Y 在解码端动态生成, 是我们研究的重点. 根据式(2), 要想使编码效率尽量高, 则式(3)中的时域相关噪声 $N(i, j, t)$ 则尽量要小, 也即 Y 与 X 要尽量相关. 图 3 显示影响 Y 的主要因素有两个: 一个是 DVC 中编码系数所在的

域, 是像素域还是变换域; 另一个是 Y 的生成方式.

3.1 变换域虚拟依赖信道模型

在变换域中, 视频信号 $X(i, j, t)$ 经过变换 R 表示为 $S(i, j, t)$,

$$S(i, j, t) = R(X(i, j, t)) \quad (4)$$

在变换域生成的边信息为

$$\begin{aligned} \tilde{S}(i, j, t) &= R(Y(i, j, t)) \\ &= R(f(X_{\text{decoded}}(i, j, t+1), X_{\text{decoded}}(i, j, t-1), \dots, \\ &\quad X_{\text{decoded}}(i, j, t-N))) \\ &= R(X(i, j, t) + N(i, j, t)) \quad (5) \end{aligned}$$

根据(4)、(5)两式, 我们可以把变换域的时域相关噪声表示为

$$\begin{aligned} N_T(i, j, t) &= S(i, j, t) - \tilde{S}(i, j, t) \\ &= R(X(i, j, t)) - R(X(i, j, t) + N(i, j, t)) \quad (6) \end{aligned}$$

式(6)表明如果能够通过变换 R 抑制空域中时域相关噪声的影响, 那么在变换域就能够对虚拟依赖信道更好地建模. 进一步根据 Girod 的工作^[13], 可预测误差的功率谱密度与运动估计的准确度之间具有如下关系:

$$\begin{aligned} \Phi_{ee}(\omega_x, \omega_y) &= \Phi_{ss}(\omega_x, \omega_y)(1 + |F(\omega_x, \omega_y)|^2) - \\ &\quad 2\Re(F(\omega_x, \omega_y)P(\omega_x, \omega_y)) + \end{aligned}$$

① 在 DVC 解码中, 经常使用在时域上前后相邻的两个关键帧生成边信息.
② 编码端的量化器会影响 $p(u|y)$ 的分布. 由于本文主要研究虚拟信道的非平稳问题, 此处暂未考虑量化器的影响.

$$\Phi_{nn}(\omega_x, \omega_y) |F(\omega_x, \omega_y)|^2 \tag{7}$$

其中 ω_x 和 ω_y 分别表示水平和垂直方向频率, Φ_{ss} 表示输入视频信号的空域功率谱, $F(\omega_x, \omega_y)$ 是闭环滤波的冲击响应, $P(\omega_x, \omega_y)$ 是运动估计误差概率密度函数的 2D 傅立叶变换, $\Phi_{nn}(\omega_x, \omega_y)$ 是运动补偿不能预测的残差信号的功率谱, $\Re(\cdot)$ 表示取复数的实部. 式(7)表明对变化缓慢的低频信号来说, 即使不太准确的运动估计也可以降低较多的预测误差; 而对高频信号来说, 运动估计必须精确, 否则由于相位的反转有可能造成信号预测误差的增加. 式(7)从另外一个角度说明了在空域中通过复杂运动估计提高虚拟信道相关度存在的困难: 原始视频信号中包含的高频分量要求非常精确的运动估计, 而在不能获取当前帧的情况下这是不可能的. 但同时该公式也指明了另外一个思路, 结合式(6), 我们得出这样的结论, 如果把视频信号分解为高频信号和低频信号, 只对低频分量采用 Wyner-Ziv 编码, 那么需要的精

确预测的高频信号被剔除, 低频信号之间形成的虚拟依赖信道模型的可靠性将得到提升, 不太精确的运动估计也可以产生较好的边信息.

3.2 小波域虚拟依赖信道模型(VCMDWT)

现在关键问题转变为如何把非平稳的视频信号转变为相对平稳的信号. 与 DCT 变换相比, 小波变换具有天然的多分辨率特点, 可以把非平稳的信号分解为一系列多尺度子带的集合, 每个子带的元素变得相对稳定^[14], 尤其是低频子带. 基于此, 本文提出了一种基于小波变换域的虚拟依赖信道模型 (Virtual dependent Channel Model based on DWT, VCMDWT), 只对小波变换的 LL 低频子带 (the coarsest subband) 采用 Wyner-Ziv 编码, 而对高频系数采用 SPIHT^[15] 方法编码. 这样 DVC 中虚拟依赖信道不再是由非平稳的空域视频信号组成, 而是由相对平稳的 LL 低频子带构成. 基于小波域的相关信道模型如图 4 所示.

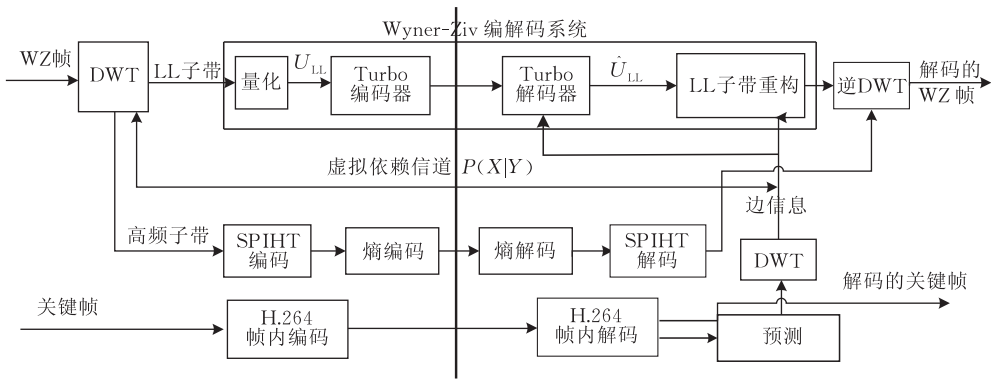


图 4 基于小波域的分布式视频编码结构及其虚拟依赖信道

为了验证所提出的基于小波域的虚拟依赖信道的有效性, 我们对空域和小波域视频信号进行了时域相关度的测试, 对视频原始信息和边信息间的相关系数进行计算, 公式如下:

$$\rho = \frac{\sum_i \sum_j (f(i, j, t) - \bar{f}(t))(s(i, j, t) - \bar{s}(t))}{\sqrt{\sum_i \sum_j (f(i, j, t) - \bar{f}(t))^2 \sum_i \sum_j (s(i, j, t) - \bar{s}(t))^2}} \tag{8}$$

其中 $f(i, j, t)$ 和 $s(i, j, t)$ 分别表示 t 时刻当前 WZ 帧和对应边信息的系数, $\bar{f}(t)$ 和 $\bar{s}(t)$ 分别表示它们所有系数的平均值. 测试序列为 Carphone 和 Foreman 前 160 个 WZ 帧. 边信息采用两种方式生成: 一种是 ME/MC, 另一种是对前后两帧的值进行平均 (average) 的方法. 测试结果如图 5 所示. 两种序列在前 80 个 WZ 帧中运动比较平缓, 像素域和小波域两种边信息生成方式下的相关系数相差不大. 图 5

中列出了运动比较剧烈的 81~160 WZ 帧的相关系数. 在 Carphone 序列中, LL 子带的相关系数数值明显高于像素域相关系数. 在小波域中两种方式生成的边信息对于相关系数值影响不大. 即使不使用 ME/MC LL 子带相关系数也要比使用 ME/MC 情况下的像素域相关系数的数值高, 尤其是包含比较剧烈运动的视频帧中, 如 90~100 和 140~160 WZ 帧两个区间. 在 Foreman 序列中, LL 子带的相关系数同样高于像素域相关系数的数值. 但是在 140~160 WZ 帧间, ME/MC 生成的边信息比平均方法生成的边信息具有更好的效果. 这是由于该区间发生了一个快速的相机平移 (Camera Pan), 也就是说在帧边界上发生了强烈运动. 即使如此, 在该区间内使用平均方法生成边信息得到的 LL 子带相关系数也与像素域使用 MEMC 情况下得到的相关系数数值相差不大.

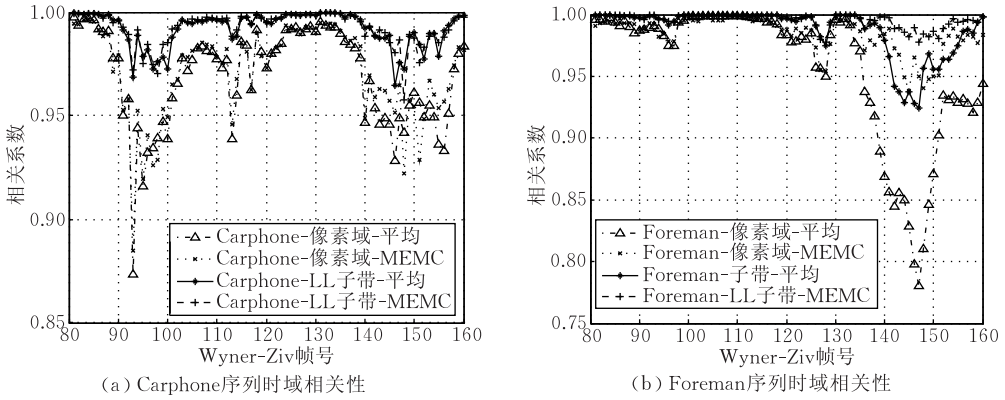


图 5 Carphone 和 Foreman 序列,在像素域和小波域,采用平均和 MEMC 两种方式生成的边信息与原始信息间的相关系数

上述测试表明了小波域低频子带间具有更高的相关性,基于低频子带构建的虚拟依赖信道具有更高的平稳度. 主要由高频信号形成的相关噪声被分离出去. 结合前面的理论分析,我们可以得出结论,如果一个变换能够把非平稳的视频信号转变或分解为较平稳的信号集,那么在该变换域建立的虚拟依赖信道相关噪声较低;边信息的生成方式对信道模型有影响,但在各域中影响程度不同.

4 实验结果

对本文提出的基于小波域的虚拟依赖信道模型的有效性在图 4 所示的 DVC 框架下进行了测试. 首先与基于像素域的方法^[4-5]进行了对比. 测试序列为 QCIF 格式的 Carphone 和 Foreman 中冲突区域较多的 80~160 WZ 帧,原始图像序列中的偶数帧作为 WZ 帧(W),奇数帧作为关键帧(I),系统编码的 GOP 结构为 I-W-I-W. 小波域的时域相关噪声也采用 Laplacian 模型建模,Carphone 和 Foreman 的 Laplacian 参数分别设为 0.17 和 0.20. 在像素域 DVC 中 Laplacian 参数采用帧级动态计算的方法^[1]

提高解码质量. DWT 变换采用 CDF97 小波,分解级数为 3. LL 子带采用 Wyner-Ziv 编码,其他子带采用 SPIHT 编码. 在像素域和小波域两种编码器中量化级均为 16 阶,使用的 turbo 码编码器相同,均由两个并行级联、码率为 4/5 的卷积码组成^[1,4]. 由于 turbo 码编码器的限制,码率最高为 2bpp,在像素域中 turbo 码的码率为 742.5kbps(因为没有使用文献[4]中的 RCPT 功能,所以码率较高),在小波域中 turbo 码的码率为 11.60kbps,为了更好地对比编码效果,SPIHT 编码码率设为 1.0bpp(编码时需要扣除 turbo 码所占码率),在经过熵编码后码率为 352.3kbps @ Carphone 和 364.5kbps @ Foreman,约为像素域码率的一半. 小波域和像素域两种编码情况下的关键帧均采用 H.264 intra 模式编解码,解码后的关键帧平均 PSNR 为 37.2dB 左右.

实验结果如图 6 所示. 在 Carphone 和 Foreman 序列中,所提方法的编码质量比基于像素域的方法,在 Carphone 中平均提高 4.5dB,在 Foreman 中平均提高 2.6dB. 在 Carphone 序列中,平均和 MEMC 两种方法生成的边信息对所提方法的解码质量影响

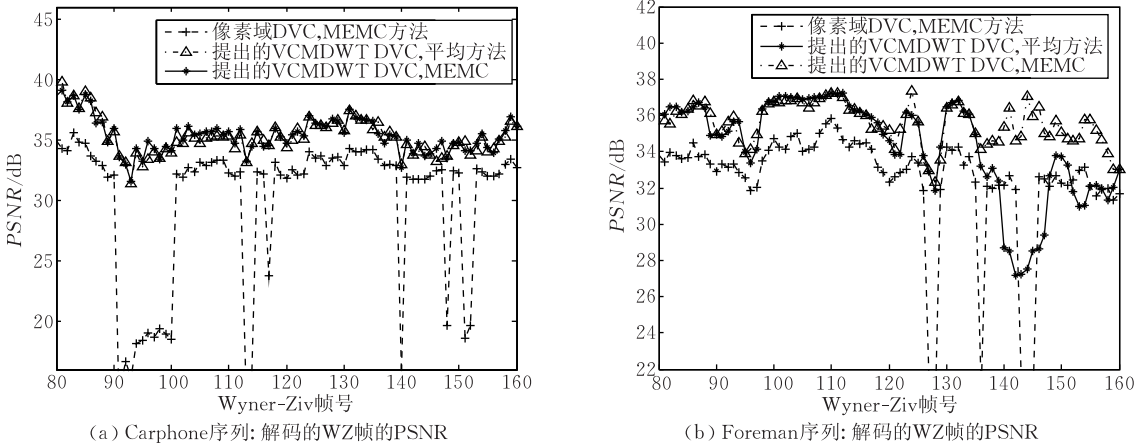


图 6 小波域与像素域虚拟依赖信道模型的编码性能对比

不大. 在 Foreman 序列中, 在 140~160 的相机 pan 运动中, MEMC 生成的边信息对于提高解码质量有较大的帮助. 图 6 显示的解码质量与图 5 中相关系数的趋势一致, 当某种方法的相关系数较高时该方法在解码时也能获得更好的解码效果. 根据实验在像素域中使用平均方法生成边信息得到的解码质量较 MEMC 方法相差较大, 本文没有列出, 具体数据可参考文献[4]. 图 6 的实验结果表明了第 3 节理论分析的有效性.

图 6 还证明了所提出的方法在运动剧烈的帧中尤其适用. 在像素域中由于图像冲突区域较多, 导致虚拟依赖信道模型不可靠而不能正确解码的帧, 在基于小波域的方法中基本都可以正确解码. 例如 Carphone 中 90~100 WZ 帧. 图 7 给出了相应的实例图. 为了更方便地对比, 我们没有直接显示 LL 子带形成的图像(因为尺寸太小), 而是把 LL 子带以其重构图像的方式显示出来, 如图 7(g)、(h)所示. 在基于 98 和 99 关键帧 MEMC 技术生成的第 99 WZ 帧的边信息图(c), 使用白色方框标示出了与原始 WZ 帧主要的不匹配区域, 这些区域基本上是

由于冲突区域的存在导致预测产生误差而生成的. 由于边信息与原始帧间的误码率较高, 像素域中的虚拟依赖信道呈现强烈的非平稳特性. 由于此时信道模型与使用的信道编码的校验能力不匹配, 导致 Wyner-Ziv 解码失败(e). 与之对照的是在所提方法中, 高频信号和低频信号分离, 对照图 7(g)和(h)可以看出低频信号间的变化比较缓慢, 因而形成的虚拟依赖信道也呈现较好的平稳特性, 这就保障了 Wyner-Ziv 解码的成功(f). 低频子带在整个图像中所占能量的比例非常高, 根据我们的统计 CDF97 小波 3 级分解后 LL 子带在所有子带中所占能量比例大于 98.5%(h). 因此在给定的编码码率中扣除 LL 子带 Wyner-Ziv 编码所需的码率后, 剩余的码率进行 SPIHT 编码, 对占据能量比例极少的少量高频系数仍然非常有效, 基本上重要的边缘均可以保留下来, 如图 7(i)和(j)所示. 因此通过把图像分解为高频信号和相对平稳的低频信号, 分别采用 SPIHT 编码和 Wyner-Ziv 编码, 可以有效地应对冲突区域造成的视频信号的非平稳特性, 同时还保障了较好的编码效率.

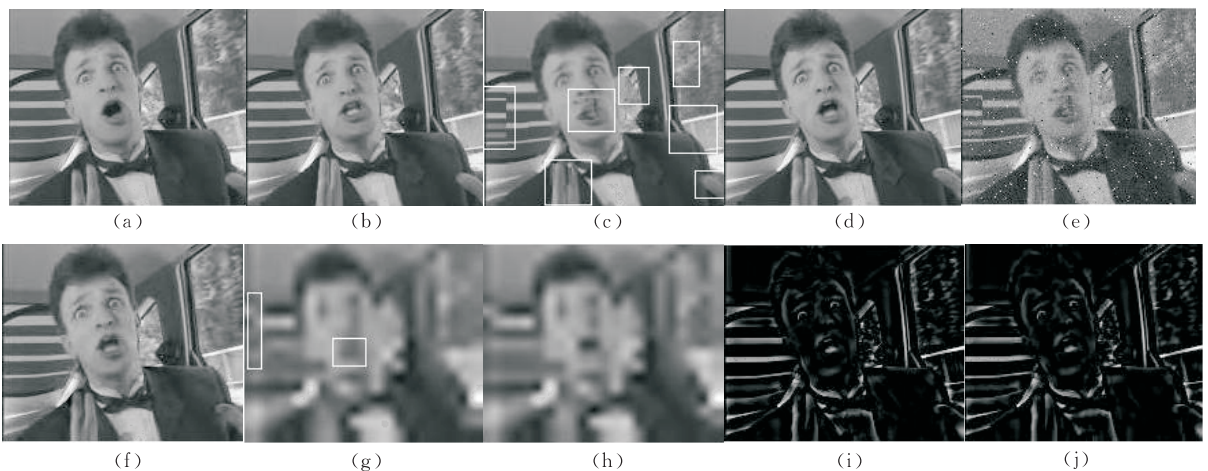


图 7 实例图(自左至右、自上至下依次为 Carphone 第 98 (a)、99 (b) 关键帧原始图像, 基于 98 和 99 关键帧使用 MEMC 技术生成的第 99 WZ 帧的边信息(c), 第 99 WZ 帧原始图像(d), 像素域解码重构的 99 WZ 帧(e), 所提方法解码重构的 99 WZ 帧(f), 由(c)的 LL 子带生成的低频分量为主的图像(g), 由(d)的 LL 子带生成的低频分量为主的图像(h), 由(d)的高频子带系数生成的图像(i)以及(i)所代表的高频系数被 SPIHT 编码后在解码端重构的图像(j))

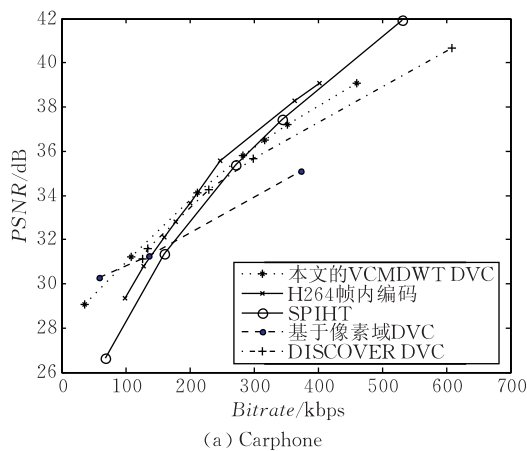
其次对提出的基于小波域虚拟依赖信道模型的 DVC 编码方法与其他编码方法在不同的码率上进行了对比, 实验结果如图 8 所示. 其中 DISCOVER DVC 编码系统工作在 DCT 域^①; SPIHT 方法采用逐帧帧内编码方式; H.264 采用帧内编码模式. 3 种 DVC 结构编码时均采用 I-W-I-W 的 GOP 结构, 帧率为 30 帧/s. 从图 8 可以看出基于像素域的 DVC 编码方法比基于 DCT 域的 DISCOVER 和本文提

出的基于小波域的 VCMDWT DVC 编码方法整体性能要低很多, 这与第 2 节的分析结果是一致的. 在图 8 中, 所提出的 VCMDWT DVC 编码方法比 DISCOVER 方法在 Carphone 和 Foreman 序列中分别高约 0.6dBs 和 0.2dBs, 并且在 Carphone 序列

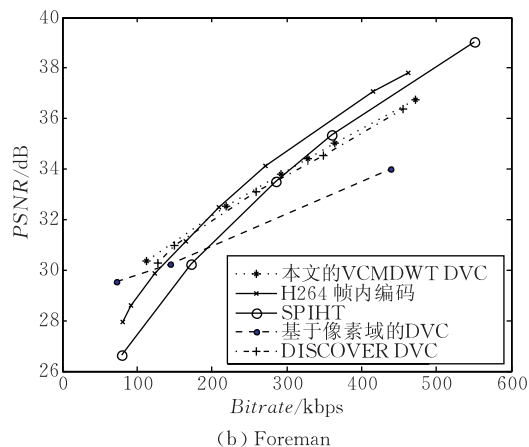
① THE DISCOVER CODEC, the IST FET programme of the European Union within the FP6/2002/IST/C Call, http://www.discoverdvc.org/cont_Codec.html, download Oct. 01, 2008

中当码率大于 352k 后, VCMDWT DVC 方法比 DISCOVER 方法 PSNR 增加更多. 根据分析, 这主要是由于在低码率时, DVC 编码涉及到的主要是低频系数的高比特位, 此时两种模型均能较好地发挥功能, 而在高码率时由于更多低频系数的低比特位以及高频系数需要编码, 此时单纯使用虚拟依赖信道模型的相关性结构进行 Wyner-Ziv 编码显然效果要差, 而所提出的 VCMDWT DVC 编码方法对高频系数采用 SPIHT 编码, 一定程度上对这一缺陷进行了弥补. 在两个测试序列中, 在码率低于约 200k 时 VCMDWT 和 DISCOVER DVC 的编码性能均比 H.264 intra 要好, 在码率低于约 300k 时均比 SPIHT 方法要好; 大于以上码率时则比 H.264 intra 和 SPIHT 方法编码性能差, 并且这种趋势随

码率的增加而增大. 这主要是由于随着码率的增大, 所有系数的低比特位和高频系数对解码质量的影响增大, 而虚拟依赖信道模型的相关性结构对低比特位和高频系数的失效逐渐增多. 通过以上实验结果可以得出结论: (1) 所提出的低频系数基于 VCMDWT 进行编码、高频系数采用 SPIHT 编码的混合 DVC 方案比基于 DCT 域的 DISCOVER 和像素域的 DVC 编码方法能够获得更好的编码性能. (2) DVC 编码技术在低码率下能够获得比基于 DCT 和小波技术的传统视频帧内编码方案 (H.264 intra 和 SPIHT) 更好的编码性能, 但是在高码率下性能要差. (3) 在高码率下, 低频系数基于 VCMDWT 进行编码、高频系数采用 SPIHT 编码的 DVC 方案有助于提高编码效率.



(a) Carphone



(b) Foreman

图 8 小波域虚拟依赖信道模型与其他编码方法的编码性能对比

5 相关工作

从 20 世纪 90 年代后期, 一些实用的分布式编码算法才开始出现. 文献[16]的作者最早发起了该项工作的研究, 对统计相关的二值高斯信源使用标量和 trellis 集构造伴随式形式的信道编码, 基于解码端边信息实现 DSC 解码. 文献[4]提出了一种基于像素域的帧内编码帧间解码系统结构, 解码端时利用边信息辅助解码, 由已经解码的前后两帧的平均值或者运动估计生成对当前 WZ 帧的一个预测, 即边信息. 随后虚拟依赖信道在 DVC 系统中的地位逐渐受到重视^[1,9,17]. 目前对于降低像素域内虚拟依赖信道模型中时域噪声的影响主要有 3 种思路. 第 1 种方法是采用更为复杂的 ME/MC 技术进行边信息的构造^[4-7]. 虽然文献[12]中声明如果对运动情况估计得非常完美, 残差帧 $N(i, j, t)$ 将呈现为真

正的白噪声, 但是即使使用非常复杂的 ME 技术, 由于冲突区域 (occlusion areas) 等情况的存在, 这个目标仍然很难达到, 而在 DVC 中由于解码端不能获取当前编码的信息, 情况更为恶化. 第 2 种思路则是采用统一的虚拟依赖信道模型, 典型的如 Laplacian 模型, 但是在解码时则动态的调整模型的参数. 例如文献[8]对像素域虚拟依赖信道模型中的 Laplacian 参数的估计方式进行了帧级、块级和像素级的研究. 第 3 种思路与第 2 种正好相反, 针对冲突区域的特性, 在建模时不再采用一个统一的模型, 而是采用混合模型, 对具有不同时域相关噪声的区域采用不同的参数建模. 典型的如把区域划分为冲突区域和非冲突区域^[1,9]. 然而这种方法仍然存在较大的困难, 主要在于冲突区域如何实现自动划分. 在不能获取当前帧信息的情况下, 这是一个非常艰巨的任务. 在对 DVC DCT 变换域的研究中, 文献[18]提出了一种基于 DCT 域的 DVC 结构, 使用伴随码

进行解码,只有 DCT 变换的低频系数采用 Wyner-Ziv 编码机制,对于高频系数要么直接丢弃,在解码端利用边信息的高频系数代替,要么直接采用帧内编码的方式;文献[19]提出了基于 turbo 码对 4×4 变换的 DCT 系数进行编码的 DVC 结构,把所有的 DCT 系数分成不同的子带分别进行 Wyner-Ziv 编码,虚拟依赖信道模型参数采用预先估计的数值,这对于低频子带是有利的,但是高频子带间形成的虚拟依赖信道仍然可靠性较低.文献[10]在文献[18]的基础上提出使用基于“小波块”的 DVC 编码结构,文献[11]提出了一种基于 SPIHT 编码技术的 DVC 编码结构,但两种方法仍然是把高频信息和低频信息统一放在一个虚拟依赖信道中处理.

6 结 论

当前分布式视频编码与传统的基于 MCP 的编码技术在性能上的差距,表明了仅仅依靠 turbo 码等信道编码的纠错能力并不能完全克服 DVC 编码中时域相关噪声的影响,这主要是由于在 DVC 中解码端不能获取当前帧足够的信息以及视频固有的非平稳特性造成的.本文对原始信息和解码端边信息组成的虚拟依赖信道进行了研究,针对像素域中虚拟依赖信道建模的困难,在变换域对虚拟依赖信道的时域相关噪声进行了分析.理论分析表明如果采用的变换能够较好地抑制空域中的时域相关噪声,那么在相应的变换域中就能够构建具有较好的平稳特性的虚拟依赖信道.根据分析结果,结合小波技术的特点,基于分离视频信号中低频和高频能量的方法,提出了一种新的基于小波变换的虚拟依赖信道模型.实验证明对于具有较多冲突区域的图像,效果尤其明显.今后将对基于小波变换域虚拟依赖信道模型的 DVC 编码机制进行更深入的研究,主要包括信道编码器和 SPIHT 编码器之间的码率分配问题、小波低频子带的量化问题和率失真问题以及对噪声进行更为详细的分类分析和处理.

参 考 文 献

- [1] Meyer A, Westerlaken P, Gunnewiek K, Lagendijk R. Distributed source coding of video with non-stationary side-information//Proceedings of the Visual Communications and Image Processing (VCIP2005). Beijing, 2005: 12-15
- [2] Slepian J, Wolf J. Noiseless coding of correlated information sources. IEEE Transactions on Information Theory, 1973, IT-19: 471-480
- [3] Wyner A, Ziv J. The rate-distortion function for source coding with side information at the decoder. IEEE Transactions on Information Theory, 1976, IT-22 (1): 1-10
- [4] Aaron A, Zhang R, Girod B. Wyner-Ziv coding of motion video//Proceedings of the Asilomar Conference on Signals and Systems. Pacific Grove, CA, 2002: 240-244
- [5] Ascenso J, Brites C, Pereira F. Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding//Proceedings of the 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services. Smolenice, Slovak Republic, 2005
- [6] Klomp S, Vatis Y, Ostermann J. Side information interpolation with sub-pel motion compensation for Wyner-Ziv decoder//Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP2006). Setúbal, Portugal, 2006
- [7] Kubasov D, Guillemot C. Mesh-based motion-compensated interpolation for side information extraction in distributed video coding//Proceedings of the International Conference on Image Processing (ICIP2006). Atlanta, USA, 2006: 261-264
- [8] Brites C. Advances on distributed video coding [Ph. D. dissertation]. Instituto Superior Tecnico, Universidade Tecnica DE Lisbon, Lisbon, 2005
- [9] Westerlaken P, Gunnewiek K, Lagendijk R. The role of the virtual channel in distributed source coding of video//Proceedings of the International Conference on Image Processing (ICIP2005). Genova, Italy, 2005: 581-584
- [10] Fowler J, Tagliasacchi M, Pesquet-Popescu B. Wavelet-based distributed source coding of video//Proceedings of the European Signal Processing Conference. Antalya, Turkey, 2005
- [11] Guo X, Lu Y, Wu F, Gao W, Li S. Wyner-Ziv video coding based on set partitioning in hierarchical tree//Proceedings of the IEEE International Conference on Image Processing (ICIP 2006). Atlanta, GA, USA, 2006: 601-604
- [12] Ishwar P, Prabhakaran V, Ramchandran K. Towards a theory for video coding using distributed compression principles//Proceedings of the 2003 International Conference on Image Processing (ICIP2003). Barcelona, Spain, 2003, 3: 687-690
- [13] Girod B. The efficiency of motion-compensating prediction for hybrid coding of video sequences. IEEE Journal of Selected Areas in Communications, 1987, 5(7): 1140-1154
- [14] Wang Y, Ostermann J, Zhang Y. Video Processing and Communications. Beijing: Prentice Hall and Tsinghua University Press, 2005
- [15] Said A, Pearlman W. A new, fast and efficient image codec based on set partitioning in hierarchical trees. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(3): 243-250

- [16] Pradhan S, Ramchandran K. Distributed source coding using syndromes (DISCUS): Design and construction//Proceedings of the IEEE Data Compression Conference. Snowbird, UT, 1999: 158-167
- [17] Pradhan S, Chou J, Ramchandran K. Duality between source coding and channel coding and its extension to the side information case. IEEE Transactions on Information Theory, 2003, 49(5): 1181-1203



FANG Sheng, born in 1971, Ph. D., associate professor. His main research interests include video coding technologies, multimedia communication, and information consumer appliances.

LI Zhe, born in 1983, M. S. candidate. His main research interests include video coding technologies and com-

- [18] Puri R, Majumdar A, Ramchandran K. PRISM: A video coding paradigm with motion estimation at the decoder. IEEE Transactions on Image Processing, 2007, 16(10): 2436-2448
- [19] Aaron A, Rane S, Setton E, Girod B. Transform-domain Wyner-Ziv codec for video//Proceedings of the SPIE Visual Communications and Image Processing. San Jose, CA, 2004: 520-528

munications.

LIANG Yong-Quan, born in 1967, Ph. D., professor, Ph.D. supervisor. His main research interests include multimedia database, artificial intelligent.

ZHONG Yu-Zhuo, born in 1938, professor, Ph. D. supervisor. His main research interests include multimedia data compression, streaming media and digital home network platform.

Background

The work of this paper is supported by National Natural Science Foundation of China (NSFC) Key Program under contract No. 90718011.

As more and more mobile devices join in the networks, especially wireless networks, encoding video anywhere any-time has become a trend. An emerging approach to the low complexity video encoding is distributed video coding (DVC, also called Wyner-Ziv video coding) with side information (SI). DVC suffers from a performance gap with respect to classical motion-compensated predictive coding solutions partly on count of the difficulty in modeling the temporal correlation noise without access to current frame. In another word, DVC decoding efficiency heavily depends on the capacity to model the temporal correlation noise between the source data and the SI, which can be regarded as the input and output of a virtual dependency channel. Although many methods have been proposed to model the virtual channel well, such as building different models corresponding to the segmentation of the video areas, make the model adaptive in frame, block or pixel level, it still remains a hard challenge and does need of further research in both theory and operation.

During the last years the authors have been carrying out research on improving the side information quality, the design of quantizer and the practical DVC system, and integra-

ting the Wyner-Ziv techniques with traditional video coding system. How to generate side information with better quality is our research focus.

This paper stemmed from the idea of classification-based video coding principle, which had been widely used in traditional video codec. However it is still in need of further research in DVC. It is well known the high frequency subbands is especially sensitive to inaccurate motion search, which is inevitable in DVC. The proposed VCMDWT model, in nature, belongs to classification-based coding paradigm, during which the LL subband with more stationary characteristics is encoded by Wyner-Ziv encoder, at the same time the non-stationary high frequency subbands are encoded by SPIHT with intra mode, in order to reduce the temporal correlation noise resulted from the high frequency components. The simulation results show that the proposed model is especially adaptable to the frames with many occlusion areas, and finally improves the rate distortion performance.

We believe the classification-based coding paradigm will also be helpful for improving rate distortion performance of DVC. In future, we will further research how to reduce the noise according to its causes, for example, the noise from the error of dense motion fields and the noise from the translation model failure.