

基于高斯混合模型的 Wyner-Ziv 视频编码

尹 明 蔡述庭 谢 云

(广东工业大学自动化学院 广州 510006)

摘 要 针对现有 Laplacian 模型不能精确描述相关噪声分布,分布式视频编码(Distributed Video Coding, DVC)的率失真性能改善非常有限,文中提出一种基于高斯混合模型的分布式视频编码方法. 首先分析了 WZ 帧与相应边信息之间相关噪声的统计特征,发现相关噪声信息的分布并不满足某种单峰分布,然后采用高斯混合模型(Gaussian Mixture Model, GMM)对噪声系数直方图进行拟合,提出基于样本特征的 EM(Expectation Maximum)算法来估计模型参数. 将提出的高斯混合相关噪声模型与相应的 Laplacian 模型进行比较,实验结果表明前者更能精确描述相关噪声的统计特征,基于该模型的 DVC 率失真性能优于基于 Laplacian 模型的 DISCOVER 方案,获得的平均增益接近 1 dB.

关键词 分布式视频编码; 相关噪声; 高斯混合分布; 无线传感器网络; EM 算法

中图法分类号 TP391 **DOI 号**: 10.3724/SP.J.1016.2012.00172

Wyner-Ziv Video Coding Based on Gaussian Mixture Model

YIN Ming CAI Shu-Ting XIE Yun

(Faculty of Automation, Guangdong University of Technology, Guangzhou 510006)

Abstract Despite recent advances, the rate-distortion (RD) performance of distributed video coding (DVC) is not yet at the level of traditional motion-compensated predictive coding. The one of critical factors is that the correlation noise model based on Laplacian cannot describe the virtual channel accurately without access to current WZ frame. To resolve this problem, the model based on Gaussian mixture model (GMM) is proposed for Wyner-Ziv video coding in this paper. First, the statistical features of the correlation noise between WZ frame and corresponding side information are analyzed. It is found that the actual probability distribution does not obey single peak distribution. We then introduce Gaussian mixture distribution to character the noise coefficient histogram. Moreover, the histogram feature is exploited to estimate the model parameter. Compared with the Laplacian model, the experimental results show that the correlation noise model based on GMM can describe the correlation noise more accurately and the performance of DVC based on Gaussian mixture model outperform the DISCOVER based on the Laplacian one with an average gain approximate to 1 dB.

Keywords distributed video coding; correlation noise; Gaussian mixture distribution; wireless sensor network; Expectation Maximum algorithm

1 引 言

随着以无线多媒体传感器网络(Wireless Mul-

timedia Sensor Network, WMSN)为代表的新应用需求日益增长,对视频编码技术提出了新的要求,即在保证较高的压缩效率的同时,要求编码复杂度较低、功耗低,解码端可以是具有较高复杂度的解

码器^[1-2]. 然而,传统的视频编码技术(如 H. 26x、MPEG 等面向广播视频应用类的方案)依赖于 DCT 变换和混合块运动补偿编码架构,其编码复杂度是解码复杂度的 5~10 倍,功耗较大,不再适用于新型应用场合. 近年来,新兴的分布式视频编码(Distributed Video Coding, DVC)以其低复杂度的编码特性、良好的编码效率,满足了对编码的计算功率、内存有严格限制的应用需求,诸如无线视频传感器网络(wireless video sensor network)、多源图像捕获系统等,引起了人们强烈的兴趣^[3-4].

DVC 建立在由 Slepian 和 Wolf 提出的分布式信源编码理论——Slepian-Wolf 理论基础之上, Wyner-Ziv(WZ) 视频编码是 DVC 的有损压缩形式. 不同于传统视频编码技术的“联合编码、联合解码”压缩方式, Wyner-Ziv 视频编码采用独立编码、联合解码的方式,可将运动估计、运动补偿等高计算复杂度的部分灵活地转移到解码端,使得编码器的复杂度大大降低^[5],从而满足了无线传感器网络应用所要求的低复杂度. 在编码效率方面,为了获得与传统编码技术近似等同的性能, Wyner-Ziv 编码在解码器端利用边信息(Side Information, SI)预测待编码帧 X,边信息由已获得的关键帧和重构帧信息通过插补方式得到^[6],被视作经过信道传输后得到的受噪声污染的待编码帧版本,而 Wyner-Ziv 编码的解码过程正是对受噪声污染后的帧纠错的过程. 由分布式信源编码理论可知,预测的边信息 SI 越接近 X,解码端用于纠错的校验信息就越少,编码器的压缩效率就越高且复杂度越低. 因此, X 与 SI 之间的密切关系表现为一种依赖关系,被视作虚拟相关信道. 这种相关性体现为原始信号与边信息的差距(即相关噪声),而相关噪声模型(Correlation Noise Model, CNM)正是表征相关性的模型,其精确与否关系到对噪声分布情况描述准确与否,很大程度上影响着分布式视频编码率失真性能^[7].

目前, CNM 模型的研究型主要集中在像素域和变换域(DCT 域和小波域). 文献[8]分析了像素域的编码码流,指出相关噪声的分布是随边信息的变化而改变,两者密切相关. Brites 等人^[9]针对像素域和变换域的 Wyner-Ziv 编码分别提出了帧、块、像素三级估计模型参数的算法,通过精确的 CNM 改善了编码效率. Slowack 等人^[10]考虑了量化噪声对 SI 帧精度的影响,并将此影响纳入到相关噪声模型的参数计算中,从而使得粗量化情况下 WZ 帧的码率最大降低了 19.5%. Cheung 等人^[11]通过抽样方

法研究了分布式信源编码的相关模型,建立了针对单二进制信源、多二进制信源及连续信源的模型,精确估计了模型中的各种参数,有效改善了编码效率. Huang 等人^[12]提出利用交叉频带信息来估计相关模型参数,可以有效改善率失真性能. Fang 等人^[13]研究了小波域中低频子带(LL subband)相关噪声信息的分布情况,并采用拉普拉斯(Laplacian)模型逼近其分布,实验表明在高码率下 RD 性能优于像素域 DVC. 在国内,房胜等人^[14]对影响时域相关噪声的因素进行了研究,提出了一种基于小波变换域的相关噪声模型 VCMDWT,提高 DVC 系统的率失真性能达到 2.6 dB 以上. 王凤琴等人^[15]发现其残差变换后直流系数带统计分布相比零均值拉普拉斯分布存在一定偏差,提出了一种基于变换域的相关噪声分布模型及自适应参数估计算法,有效提高了编码效率.

然而,上述研究工作中均假定相关噪声分布服从 Laplacian 模型,工作大多集中在参数动态估计方面. 虽然 Maugey 等人^[16]采用幂指数模型(exponential power distribution)来拟合相关噪声,仿真表明编解码系统性能优于 Laplacian 模型,但没给出全面反映新模型的精度及率失真性能改善的实验结果. 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,相关噪声呈现非平稳特性. 尽管 Laplacian 模型易于计算,但用于描述相关噪声概率分布并不精确,不能反映所有的视频运动模式可能性,从而增加 WZ 帧解码所需的比特数,导致 Wyner-Ziv 编码系统的效率损失,这正是其率失真性能难以到达传统运动补偿预测编码性能的症结之一^[17]. 因此,本文深入研究了不同运动类型视频序列的原始帧与相应边信息之间噪声系数的分布情况,发现相关噪声信息的分布并不满足某种单峰分布,并利用高斯混合模型(Gaussian Mixture Model, GMM)理论^[18],提出了一种新的 DCT 域分布式视频编码系统的相关噪声模型. 针对模型参数估计问题,提出了基于样本特征的期望最大(expectation Maximum, EM)算法.

2 相关工作

2.1 DCT 域 Wyner-Ziv 编解码体系结构

变换域 Wyner-Ziv 视频编码(TDWZ)是目前典型的分布式视频编码方案^[4],图 1 给出了本文采用的 DCT 域 Wyner-Ziv 视频编码架构. 它由量化器和 Slepian-Wolf 编码器(SWC)构成. 在编码端,视频图

像帧通常被分为两类: 关键帧(即 Key 帧)和 WZ 帧. Key 帧采用传统帧内编码方法进行编解码, 而 WZ 帧采用 Wyner-Ziv 编码. 具体说来, Wyner-Ziv 编解码系统的工作原理为: 首先对序列 X 进行变换处理(如 DCT 变换)以去除图像空域相关性, 然后对相应变换系数进行量化, 产生量化符号流. 量化后的符号流经由 Slepian-Wolf 编码器产生量化符号流的奇偶校验信息. 在解码端, 利用已解码帧进行边信息估计, 得到边信息帧 SI, 结合 Slepian-Wolf 解码出的校验比特一起重构出 WZ 帧. 一旦解码质量未达到误码率的要求时, Wyner-Ziv 解码端会通过反向信道要求 Slepian-Wolf 编码器提供更多的校验信息. 因此, 当边信息 SI 越接近原始 WZ 帧, 解码端用

于纠错的所需校验信息就越少, 这反映了边信息 SI 与原始 WZ 帧存在紧密的相关依赖. 为了更好地利用边信息解码和重建 WZ 帧, 相关噪声的分布情况变得非常重要. 它不仅反映了虚拟信道的特性, 还影响着解码端 WZ 帧的重构质量. 但实际编解码过程中, 原始 WZ 帧和 SI 分别出现在编、解码端, 解码端无法获知原始信息, 而编码端也只能以离线方式重构边信息或通过反向信道从解码端传输, 进而增加系统计算复杂度及时延等诸多不利, 这不符合实际应用需求. 而且随着编码视频内容的变化, SI 在时域和空域上动态变化, 相关噪声亦呈现非平稳特性, 尤其对于运动较复杂的序列边信息质量变差, 因此对相关噪声的分布精确建模变得非常困难.

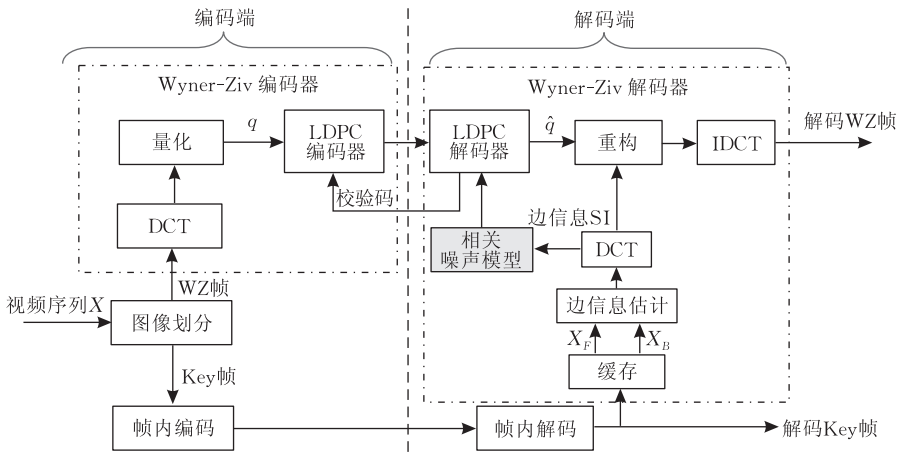


图 1 DCT 域 Wyner-Ziv 编解码系统架构

目前, 常用的边信息生成方法是基于运动补偿内插算法 (Motion Compensation Interpolation, MCI), 其操作过程为: 首先根据图像组结构确定用于生成边信息的参考帧(即前向参考帧和后向参考帧), 通过运动搜索得到初始运动矢量信息, 然后根据双向运动估计得到最佳运动矢量, 通过交叠块运动补偿 (Overlapped Block Motion Compensation, OB-MC) 方法生成边信息. 利用式(1)求得最佳运动矢量:

$$\langle d\hat{x}, d\hat{y} \rangle = \arg \min_{d\hat{x}, d\hat{y}} \left[\sum_{i,j \in \Omega} |X_F(i+d\hat{x}_F, j+d\hat{y}_F) - X_B(i+d\hat{x}_B, j+d\hat{y}_B)|^2 \right],$$

$$(d\hat{x}_B, d\hat{y}_B) = -(d\hat{x}_F, d\hat{y}_F) = (dx, dy)/2 \quad (1)$$

其中 (i, j) 表示像素的位置, X_F 和 X_B 分别为待解码帧的前后两帧(已解码), Ω 表示为当前块的像素集合.

由此, 获取前向和后向的最佳运动矢量 $(d\hat{x}_B, d\hat{y}_B)$ 、 $(d\hat{x}_F, d\hat{y}_F)$, 进而生成边信息 $S(i, j)$:

$$S(i, j) = \frac{1}{2} X_B(i+d\hat{x}_B, j+d\hat{y}_B) +$$

$$\frac{1}{2} X_F(i+d\hat{x}_F, j+d\hat{y}_F) = X(i, j) + N(i, j) \quad (2)$$

其中, $N(i, j)$ 表示相关噪声.

经过 DCT 变换后的边信息, 一方面用于 WZ 帧解码, 另一方面用于相关噪声建模. 由式(1)可看出, 如果前后运动补偿帧相差越大, 所生成的边信息质量越差, 则相关噪声越大. 反之, 边信息质量越好, 则相关噪声越小. 实际编码过程中边信息不会出现在编码端, 解码端又无法获取原始帧信息, 所以前后运动补偿帧之差往往被采用来描述相关噪声的概率分布^[8-10]. 由式(2)得相关噪声 $N(i, j)$:

$$N(i, j) = S(i, j) - X(i, j) = \frac{1}{2} X_B(i+d\hat{x}_B, j+d\hat{y}_B) - X(i, j) + \frac{1}{2} X_F(i+d\hat{x}_F, j+d\hat{y}_F) \quad (3)$$

在本文中, 利用前后运动补偿帧之差 $R(i, j)$ 来对相关噪声建模:

$$R(i, j) = |X_B(i + dx_B, j + dy_B) - X_F(i + dx_F, j + dy_F)| \quad (4)$$

2.2 图像系数的统计模型

不少文献^[7-10]对 DCT 系数的统计特性研究表明,其概率分布近似服从 Laplacian 分布,如式(5)所示

$$f(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}}{\sigma}|x-\mu|\right) \quad (5)$$

其中, σ^2 为方差, μ 为均值。

然而,文献^[19]研究表明, Laplacian 模型并不是 DCT 系数概率分布的最佳逼近。文献^[20]论证了 Laplacian 分布具有指数级的衰减,不适合具有重拖尾的样本数据。其它密度分布函数,诸如:广义高斯分布、高斯分布、伽玛(Gamma)分布、瑞利分布以及 Laplacian 分布都往往要求所拟合系数的概率分布具有单峰形式,即只有一个极大值。但实际中,不同的图像区域 DCT 系数可能包含多峰的概率形式,在特征空间中往往表现为多种分布的混合,很难把这种复杂的分布通过单一的参数化密度函数表示出来。而高斯混合模型(GMM)在拟合多种概率密度分布情况时,具有理论分析的可解析性,表现为较强的逼近能力^[21]。因此,本文将采用高斯混合模型来逼近 DCT 系数的分布,进而精确描述相关噪声信息。

单变量的高斯随机函数:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad (6)$$

多变量的高斯随机函数:

$$g(\mathbf{X}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{X}-\mathbf{U})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{U})\right) \quad (7)$$

其中, $\boldsymbol{\Sigma}$ 为协方差矩阵, \mathbf{U} 为均值矢量, \mathbf{X} 为变量矢量, d 表示高斯分布的维数。

图像的高斯混合模型是不同的高斯函数的线性组合,其表达形式为

$$f(x) = \sum_{m=1}^M \rho^m g^m(x), \quad \sum_{m=1}^M \rho^m = 1 \quad (8)$$

其中, $g(\cdot)$ 为高斯分布函数, ρ^m 为第 m 维高斯分布的权值, M 表示模型中高斯分布的总数量。

混合高斯模型的参数集 $\Theta = (\rho, \mu, \sigma)$ 可以使用期望最大(EM)算法估计得到。EM 算法基本思想是利用训练数据 $X = (x_1, x_2, \dots, x_n)$, 从模型参数集的初始值 Θ_0 开始, 估计一个新的参数集 $\tilde{\Theta}$, 使得新的

模型参数下的似然概率 $P(X|\tilde{\Theta}) \geq P(X|\Theta_0)$ 。新的模型参数再作为当前参数进行训练, 这样迭代运算直到模型收敛。

3 相关噪声信息统计特征分析及建模

本文对几种典型的视频序列的 DCT 残差系数的统计特征进行分析, 重点比较 Laplacian 模型和高斯混合模型在 DCT 频带残差系数(亮度分量)分布上的拟合程度。在 DCT 域 Wyner-Ziv 编码系统中, 相关噪声是边信息 SI 与原始 WZ 帧之差, 其系数正是 DCT 残差系数。随着图像区域的改变, 相关噪声呈现动态变化, 其方差被视为一随机变量。

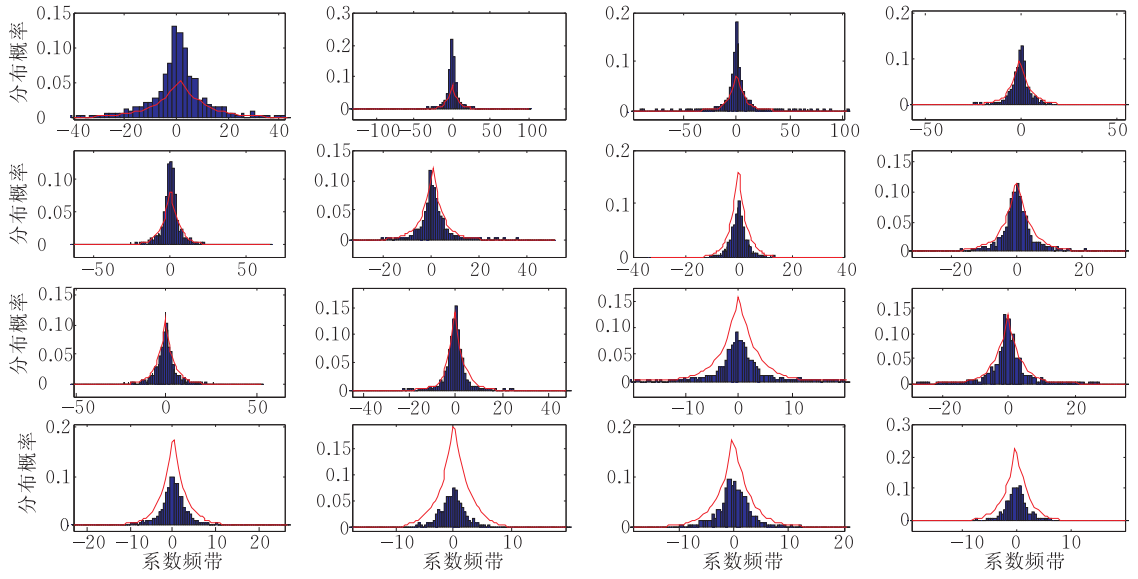
这里使用斯坦福大学提供的仿真平台^[19]来获取相关噪声数据。在仿真实验中, WZ 帧采用规则 LDPC 作为 Slepian-Wolf 编码, 符号节点的度为 3, 信息长度为 6636 比特。DCT 变换采用 8×8 块, 量化参数 QP 的选取尽量满足于使每一 Key 帧和 WZ 帧的 PSNR 值近似相等, 保证重构出的图像近似平滑。实验选用了不同运动类型的标准视频序列 Foreman、Football、Coastguard 和 Carphone, 以 QCIF 格式和 15 Hz 帧率进行编码, 共 100 帧。

从实际应用角度考虑, 在线建模相关噪声信息无法获取原始图像帧。与文献^[8-12, 15]一致, 本文利用前后运动补偿帧之差 $R(i, j)$ 来对相关噪声建模。对式(4)所得帧差 $R(i, j)$ 进行 4×4 DCT 变换后, 将频带系数分组形成频带组 R_k ($k = 0, 1, \dots, 15$), 然后采用 Laplacian 模型拟合其分布情况。图 2 给出几种典型视频序列的补偿帧残差系数分布特征, 每种序列采用所有的 WZ 帧进行统计。

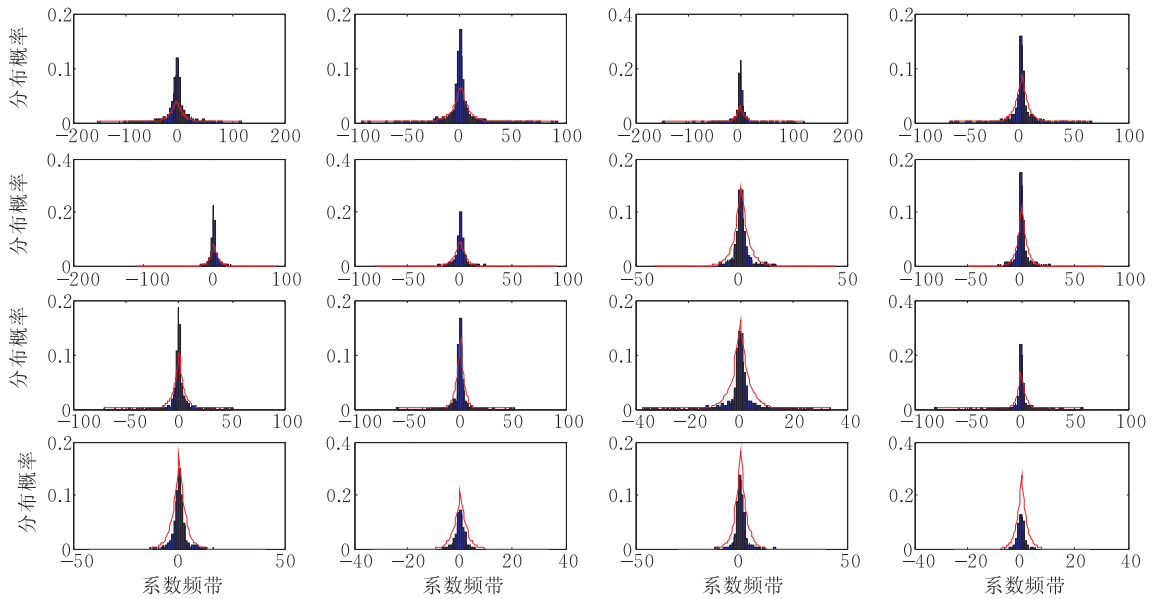
从图 2 可发现, 在不同运动类型的视频序列中, 系数频带中直流系数的均值不为零, 方差较大, 表明集中了图像大部分能量。而低频带均值近似为零, 方差稍小, 随着频率增加, 系数分布越集中, 方差越小, 但分布与近似零均值 Laplacian 分布存在一定差距, 尤其在尾部落降部分, 出现了较严重拖尾, 曲线下降较慢。

因此, 本文改进现有模型, 直流频带系数近似服从 Laplacian 分布(均值不为零), 其余频带系数采用 GMM 分布拟合, 即

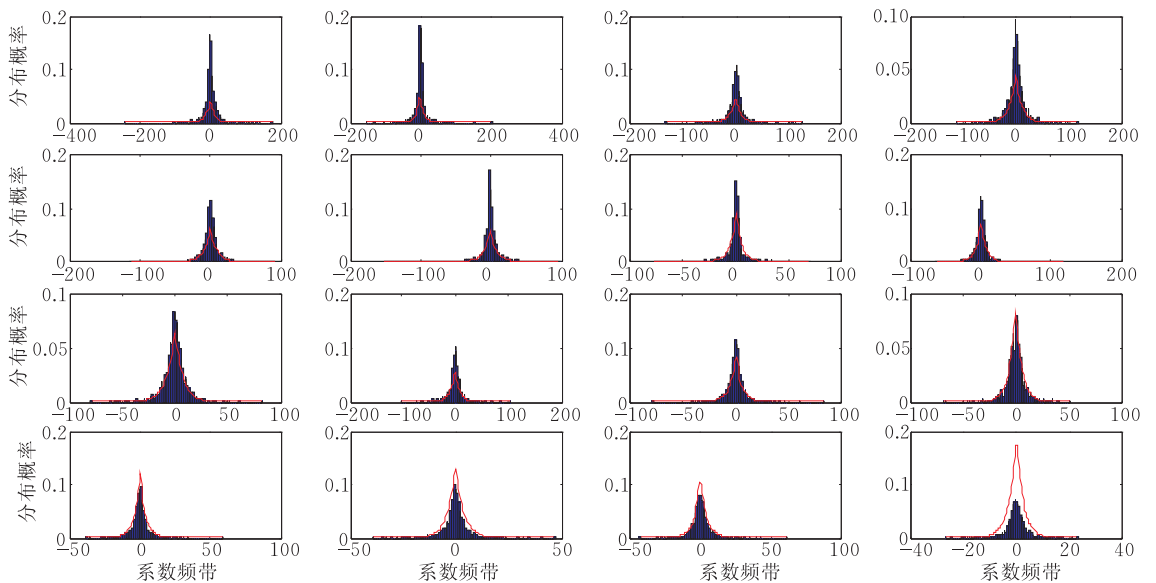
$$f(R_k) = \begin{cases} \frac{\lambda}{2} e^{-\lambda|R_k-\mu|}, & k=0 \\ \sum_m \rho^m g_{R_k}^m(\Theta), & k=1, 2, \dots, 15 \end{cases} \quad (9)$$



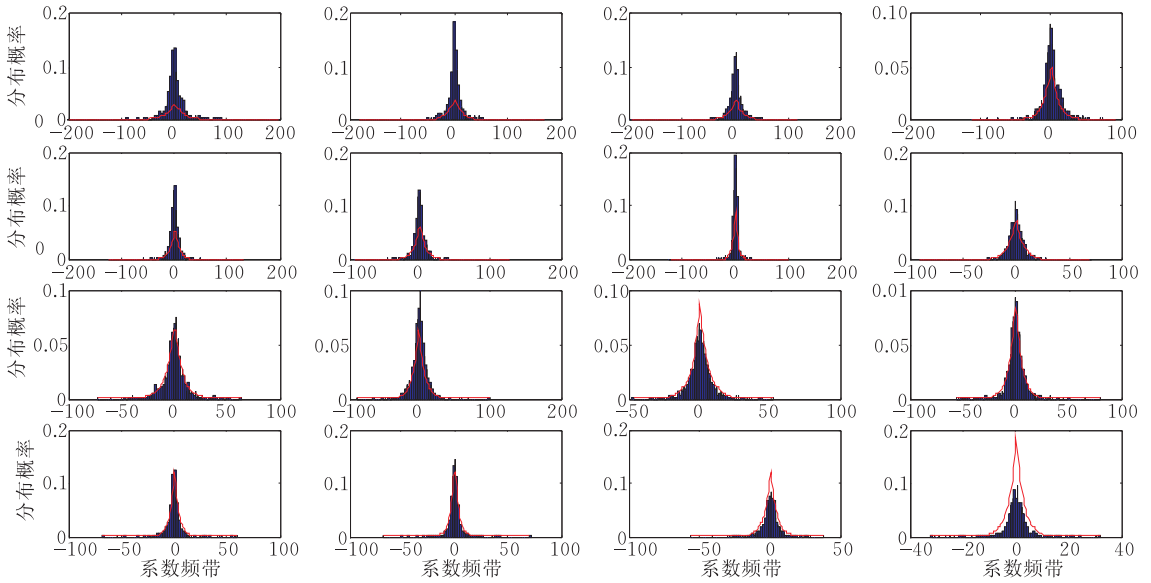
(a) Foreman序列的残差频带系数统计分布



(b) Carphone序列的频带系数统计分布



(c) Coastguard序列的频带系数统计分布



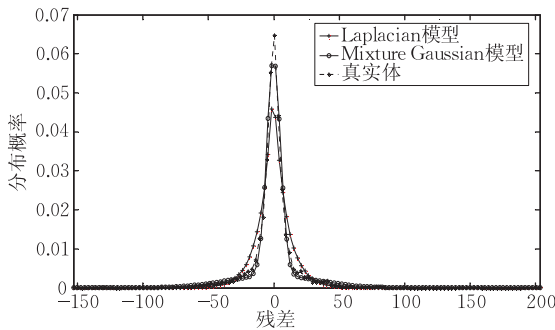
(d) Football序列的频带系数统计分布

图 2 几种典型视频序列的 DCT 域残差系数分布特征(自左向右,自上向下分别为 DC, AC1, AC2, ..., AC15 频带分布)

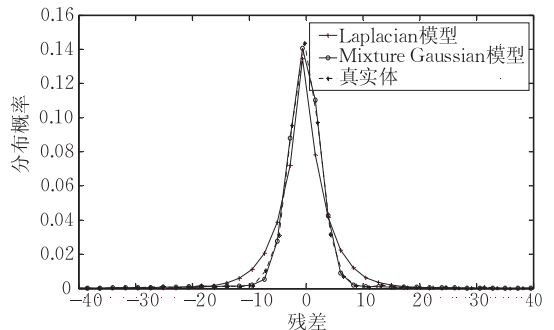
其中, $\lambda = \sqrt{2}/\sigma$, $\mu = \frac{1}{N} \sum R_0$ 为直流系数的平均值, m 表示高斯分布数量, ρ^m 为每一高斯分布的权值, 表示频带系数归属于不同的运动类型. $\Theta = \{\mu^m, \sigma^m\}$ 为高斯分布参数集.

一般来说, 不同运动类型的图像帧所包含的运动量也会不同, 当图像表现为运动剧烈时, 则解码端

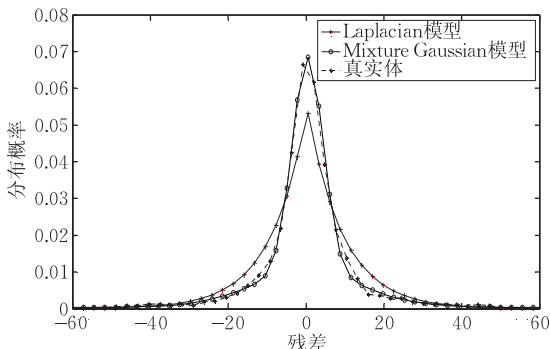
前后向补偿帧的绝对误差较大, 即残差系数所含能量就越高, 在分布上表现为大方差系数. 反之, 当图像表现为运动平缓时, 则解码端前后向补偿帧的绝对误差较小, 即残差系数所含能量就越低, 在分布上表现为小方差系数. 为计算简便, 本文假设噪声系数的方差可分作大小两类, 即大小方差分别为 σ_1^2 、 σ_0^2 , 则高斯混合相关噪声模型表示为



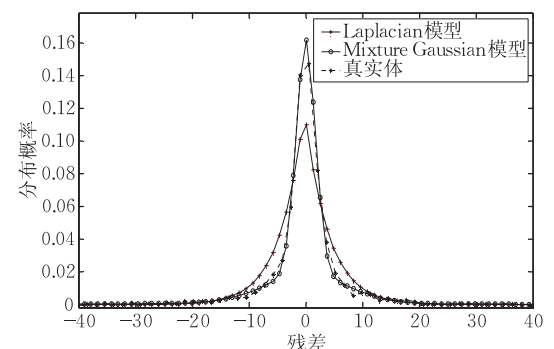
(a) Coastguard序列的频带系数统计特征(AC第1带)



(b) Hall Monitor序列的频带系数统计特征(AC第2带)



(c) Soccer序列的频带系数统计特征(DC带)



(d) Foreman序列的频带系数统计特征(AC第3带)

图 3 2 种模型对 DCT 域残差系数分布拟合的比较

$$\begin{aligned}
f(x) &= \rho^0 g^0(x) + \rho^1 g^1(x) \\
&= \rho^0 \frac{1}{\sqrt{2\pi}\sigma_0} \exp(-1/2\sigma_0^2) + \rho^1 \frac{1}{\sqrt{2\pi}\sigma_1} \exp(-1/2\sigma_1^2)
\end{aligned} \quad (10)$$

我们采用高斯混合相关噪声模型对残差系数进行了拟合,发现在低频分量部分,本文模型比 Laplacian 模型能更精确描述残差系数的分布. 由于篇幅限制,仅给出部分代表性结果,如图 3 所示.

图 3 给出了 4 种序列残差系数在低、中频带上的拟合情况. 可看出,除均值附近的几个尖锐点外, GMM 模型与系数真实分布的拟合程度非常好,出现概率较大的高斯分布方差小,代表了大量的小系数. 而出现概率较小的高斯分布方差大,代表了少量的大系数,这完全符合 DCT 残差系数的特征. 相比 Laplacian 模型, GMM 模型在尾部下降部分能更精确地拟合噪声系数的分布情况,消除了 Laplacian 模型尾部下降过慢的现象,减少了解码所需的校验数据.

4 高斯混合相关噪声模型的 EM 算法

在噪声分布建模中,残差频带系数构成观察空间,为不完备数据,而频带系数所描述的图像运动类型构成不可观察空间,为隐藏数据,两者构成完备数据可表示为 $F = \{(R_k, \rho), k \in (1, 2, \dots, 15)\}$. 假设每个频带中样本数为 N , 高斯分布数为 $M=2$, 构造似然函数 $L(\Theta; R, \rho)$:

$$L(\Theta; R, \rho) = \prod_{i=1}^N \sum_{m=1}^2 \rho^{(m)} g(R_{k,i}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (11)$$

其中, $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}$ 为协方差矩阵, $R_{k,i}$ 为第 k 频带中的第 i 个系数.

系数归属不同运动类型为 $m (m=1, 2)$ 类分布出现的后验概率为

$$\begin{aligned}
\tau_{m,i}^{(t)} &:= P(\rho_i = m | R_{k,i}; \Theta^{(t)}) \\
&= \frac{\rho_m^{(t)} g(R_{k,i}; \boldsymbol{\mu}_m^{(t)}, \boldsymbol{\Sigma}_m^{(t)})}{\sum_{j=1}^2 \rho_j^{(t)} g(R_{k,i}; \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)})}
\end{aligned} \quad (12)$$

则期望(E)步骤的目标函数如下式所示(具体推导见文献[22]):

$$\begin{aligned}
Q(\Theta | \Theta^{(t)}) &= E[\ln L(\Theta; R, \rho)] \\
&= \sum_{i=1}^N \sum_{m=1}^2 \tau_{m,i}^{(t)} \left[\ln \rho_m - \frac{1}{2} \ln |\boldsymbol{\Sigma}_m| - \frac{1}{2} \ln(2\pi) - \frac{1}{2} (R_{k,i} - \boldsymbol{\mu}_m)^T \boldsymbol{\Sigma}_m^{-1} (R_{k,i} - \boldsymbol{\mu}_m) \right]
\end{aligned} \quad (13)$$

最大化(M)步骤选择 $\Theta^{(t+1)}$,

$$\Theta^{(t+1)} = \arg \max_{\Theta} Q(\Theta | \Theta^{(t)}) \quad (14)$$

根据拉格朗日极值求解方法,得到参数求解的迭代公式:

$$\rho_m^{(t+1)} = \frac{1}{N} \sum_{i=1}^N \tau_{m,i}^{(t)} \quad (15)$$

$$\boldsymbol{\mu}_m^{(t+1)} = \frac{1}{N \rho_m^{(t+1)}} \sum_{i=1}^N R_{k,i} \tau_{m,i}^{(t)} \quad (16)$$

$$\boldsymbol{\Sigma}_m^{(t+1)} = \frac{1}{N \rho_m^{(t+1)}} \sum_{i=1}^N \tau_{m,i}^{(t)} (R_{k,i} - \boldsymbol{\mu}_m^{(t+1)}) (R_{k,i} - \boldsymbol{\mu}_m^{(t+1)})^T \quad (17)$$

上述方法是 EM 的标准算法,然而该算法收敛过慢,需要迭代多次,参数才能稳定,计算复杂度较大. 因此,本文将直方图中对应频数不为零的每个系数(四舍五入整数化后)作为一个特征,利用残差的各个频带系数特征快速迭代求解模型参数,具体算法步骤如下:

1. 统计各个频带的系数,得到相应的特征总数 N' 、每一特征值 $R_{k,1}, R_{k,2}, \dots, R_{k,N'}$ 及相应出现频数 $h_1, h_2, \dots, h_{N'}$.
2. 修正式(15)~(17),得

$$\rho_m^{(t+1)} = \frac{1}{N'} \sum_{i=1}^{N'} h_i \tau_{m,i}^{(t)} \quad (18)$$

$$\boldsymbol{\mu}_m^{(t+1)} = \frac{1}{N' \rho_m^{(t+1)}} \sum_{i=1}^{N'} R_{k,i} \tau_{m,i}^{(t)} \quad (19)$$

$$\boldsymbol{\Sigma}_m^{(t+1)} = \frac{1}{N' \rho_m^{(t+1)}} \sum_{i=1}^{N'} \tau_{m,i}^{(t)} (R_{k,i} - \boldsymbol{\mu}_m^{(t+1)}) (R_{k,i} - \boldsymbol{\mu}_m^{(t+1)})^T \quad (20)$$

基于特征的 EM 算法减少了样本数 ($N' \leq N$), 从而避免了重复计算,加速了迭代收敛,降低计算复杂度. 同时,由于减少的样本均为零值系数,所以并不影响参数的最大似然估计精度.

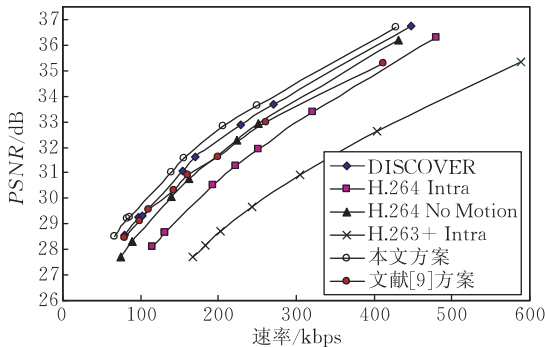
5 仿真实验及分析

为评估提出的相关噪声模型性能,本文选择了 4 组典型的视频序列进行测试,分别为 Foreman 及 Soccer、Coastguard、Hall Monitor, 对应高、中及低运动程度的视频序列. 序列均采用 QCIF@15Hz 格式, GOP 长度为 2. 根据大多文献的评估方法,实验中仅评估亮度分量的率失真结果. Key 帧采用 H. 264/AVC 帧内编码(Main profile), 依据表 1 所示量化参数量化,得到相应质量(RD)等级. 边信息采用帧内插方式[23]生成,前后向运动估计搜索范围取 ± 32 个像素,运动补偿采用 OBMC. 目前 RD 性能最好的 TDWZ 系统是 DISCOVER 方案[24], 而相

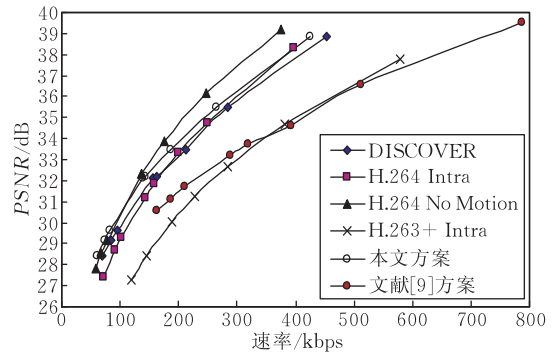
关噪声建模最好的方法为文献[9]所提出,其中相关噪声模型均采用 Laplacian 分布描述,因此本文将改进模型的性能与两者进行比较,作为性能评估.实验参考了文献[25]提供的部分源代码,采用 C 和 Matlab 混合编程.另外,实验中还选取了 3 种标准的编码方案 H.263 + Intra, H.264/AVC Intra, H.264/AVC Inter No Motion 进行率失真性能比较,用以说明与最好的传统视频编码技术性能的优劣.最后,本节比较了采用改进模型带来的计算复杂度.

表 1 对应不同质量水平的 Key 帧的量化参数

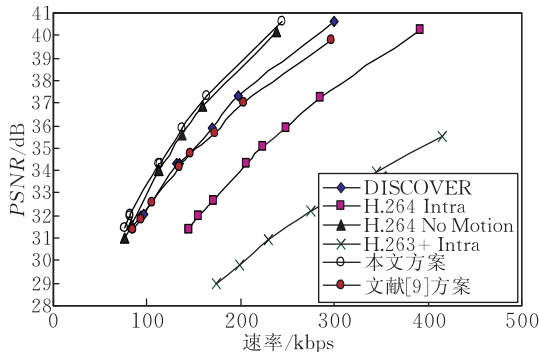
序列	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Foreman	42	40	39	36	35	33	31	26
Coastguard	39	38	38	35	34	33	31	27
Hall Monitor	37	36	35	33	32	31	29	25
Soccer	45	44	42	38	38	35	31	26



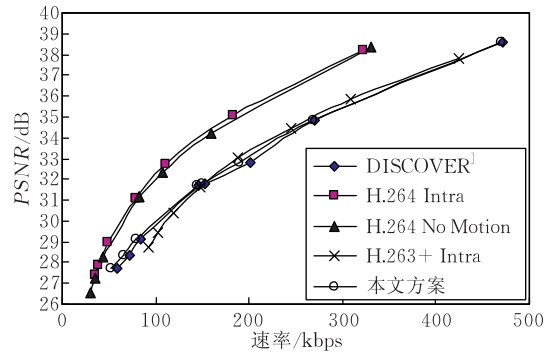
(a) Coastguard 序列的率失真性能比较



(b) Foreman 序列的率失真性能比较



(c) Hall Monitor 序列的率失真性能比较



(d) Soccer 序列的率失真性能比较

图 4 几种典型视频序列的率失真曲线(GOP=2)

随着码率的增加,噪声模型对 DVC 性能的增益越大,但对于高运动序列来说,由于高码率减轻了解码对模型精度的依赖,增益反而减小.对比其它 3 种标准编码方案,对于低运动序列来说,其性能可以与 H.264/AVC Inter No Motion 相当,超过帧内(Intra)编码的方案.但对于高运动和复杂的视频来说,所提出的 Wyner-Ziv 编码在率失真性能方面仍存在较大改善空间,尤其是包含剧烈运动的 Soccer 序列.

5.2 模型复杂度分析

模型精度是其性能的一方面,而计算模型参数

5.1 模型 RD 性能分析

从图 4 所示的率失真性能曲线,相比 Laplacian 模型,本文提出的高斯混合相关噪声模型能更精确地描述相关噪声,基于该模型的 DVC 率失真性能优于 DISCOVER 方案^[24]和文献[9]的方案,其中平均增益接近 1 dB,对于 Hall Monitor 序列最大增益可达到 1.48 dB.对于运动平缓的图像序列(如 Hall Monitor),由于解码端前后向补偿帧的绝对误差较小,利用残差系数训练的高斯混合相关噪声模型更精确地反映了虚拟信道的模型,在同等质量下,用于纠错的比特数较少,解码端获得了更高的增益.而对于运动较剧烈的图像序列(如 Soccer),解码端前后向补偿帧的绝对误差较大,无法有效建模,所需纠错比特数增加,改进模型获得的率失真增益较小.

的复杂度也关乎于改进编解码算法在实际应用中的效果.在解码端,本文利用基于样本特征的 EM 算法估计 GMM 模型参数,由于 EM 算法是一种最大似然参数估计方法,需通过多次迭代运算来实现最优参数估计,其中 GMM 模型的维数和样本集的大小是影响计算的复杂度的两个重要因素,因此,将其引入到 Wyner-Ziv 编解码系统必然引起运算复杂度的增加,本节将讨论改进模型方案与 DCT 域 Wyner-Ziv 解码的运算复杂度不同.

解码使用 3.4GHz 主频和 2.048GB RAM 内存

的 PC 机,代码的软件环境为 Visual Studio 2008,其它条件与性能评估时相同.表 2 给出了以运算时间为准则的解码复杂度对比.可以看出,利用 EM 算法估计 GMM 模型参数确实引入了较大的计算复杂度,但 Wyner-Ziv 编码架构的主要目标是寻求“编

码简单、解码复杂”,所适合的应用也允许这种情况(例如,Wyner-Ziv 系统的解码端往往是具有强大计算性能的中心节点)出现.然而,作为改进编码方案降低计算复杂度是性能改善所追求之一,因此,寻求低复杂度的优化 GMM 模型参数估计是本文后续的工作.

表 2 不同方法解码时间比较

(单位:s)

RD 点	Coastguard		Hall Monitor		Soccer		Foreman	
	DISCOVER	本文方法	DISCOVER	本文方法	DISCOVER	本文方法	DISCOVER	本文方法
Q1	295.15	1284.62	240.67	1238.10	723.66	1732.80	508.56	1563.1
Q2	408.77	1411.36	322.45	1322.50	897.62	1899.02	634.59	1602.3
Q3	475.67	1475.01	371.00	1372.20	1010.78	2021.64	709.72	1700.0
Q4	794.65	1799.32	564.22	1564.87	1651.63	2651.21	1178.80	2183.6
Q5	820.48	1817.32	605.84	1605.00	1771.77	2699.30	1290.30	2299.0
Q6	1257.99	2259.03	869.16	1872.20	2316.64	3362.00	1802.90	2803.1
Q7	1668.90	2675.99	1055.60	2101.30	2949.72	3801.98	2328.30	3327.8
Q8	3090.83	4102.12	1712.80	2799.01	4775.63	5799.20	3873.80	4872.9

6 总 结

针对分布式视频编码中相关噪声的分布并不满足某个单一密度分布,现有 Laplacian 模型不能精确描述相关噪声的分布,本文利用高斯混合模型(Gaussian Mixture Model, GMM)理论,提出了一种新的 DCT 域分布式视频编码系统的相关噪声模型.同时,针对 EM 算法估计模型参数存在收敛过慢的问题,提出了基于样本特征的改进 EM 算法.实验结果表明,相比 Laplacian 模型,本文提出的高斯混合相关噪声模型能更精确地描述相关噪声,基于该模型的 Wyner-Ziv 编码率失真性能优于 DISCOVER 方案.然而,对于包含大运动和复杂运动的视频来说,本文提出的编码方案在率失真性能方面,相比 H.264/AVC Inter No Motion,仍存在较大改善空间.此外,寻求低复杂度的优化 GMM 模型参数估计也是本文需进一步的工作.

参 考 文 献

- [1] Akyildiz I F, Melodia T, Chowdhury K R. A survey on wireless multimedia sensor networks. *Computer Networks*, 2007, 51(1): 921-960
- [2] Peixoto E, de Queiroz R L, Mukherjee D. A Wyner-Ziv video transcoder. *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, 20(2): 1453-1461
- [3] Girod B, Aaron A, Rane S, Rebollo-Monedero D. Distributed video coding. *Proceedings of the IEEE Special Issue on Advances in Video Coding and Delivery*, 2005, 93(1): 71-83
- [4] 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, California, USA, 2004: 520-528

- [5] Cai Shu-Ting, Wang Qin-Ruo, Xie Yun, Yin Ming. A kind of distributed video compression scheme based on rateless low-density parity-check codes. *Acta Electronica Sinica*, 2009, 37(11): 2535-2540(in Chinese)
(蔡述庭, 王钦若, 谢云, 尹明. 一种基于无比率 LDPC 编码的分布式视频压缩机制. *电子学报*, 2009, 37(11): 2535-2540)
- [6] Martins R, Brites C, Ascenso J, Pereira F. Statistical motion learning for improved transform domain Wyner-Ziv video coding. *IET Image Processing*, 2010, 4(1): 28-41
- [7] Westerlaken R P, Gunnewiek R K, Lagendijk R L. The role of the virtual channel in distributed source coding of video//*Proceedings of IEEE ICIP*. Genoa, Italy, 2005: 581-584
- [8] Deligiannis N, Munteanu A, Clerckx T et al. Correlation channel estimation in pixel-domain distributed video coding//*Proceedings of the WIAMIS 2009*. London, UK, 2009: 93-96
- [9] Brites C, Pereira F. Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2008, 18(9): 1117-1190
- [10] Slowack J, Mys S, Skorupa J et al. Accounting for quantization noise in online correlation noise estimation for distributed video coding//*Proceedings of the PCS2009*. Chicago, Illinois, USA, 2009: 1-4
- [11] Cheung N M, Wang H S, Ortega A. Sampling-based correlation estimation for distributed source coding under rate and complexity constraints. *IEEE Transactions on Image Processing*, 2008, 17(11): 2122-2137
- [12] Huang X, Forchhammer S. Improved virtual channel noise model for transform domain Wyner-Ziv video coding//*Proceedings of the ICASSP*. Taipei, China, 2009: 921-924
- [13] Fang S, Li Z, Zhang L W. Distributed video codec modelling correlation noise in wavelet coarsest subband. *Electronics Letters*, 2007, 43(23): 1266-1267
- [14] Fang Sheng, Li Zhe, Liang Yong-Quan, Zhong Yu-Zhuo. Research of the virtual dependency channel in distributed video coding. *Chinese Journal of Computers*, 2009, 32(7): 1404-1412(in Chinese)
(房胜, 李哲, 梁永全, 钟玉琢. 分布式视频编码虚拟依赖信道模型研究. *计算机学报*, 2009, 32(7): 1404-1412)

- [15] Wang Feng-Qin, Fan Yang-Yu, Zhao Jiong, Liu Yuan-Kui. Correlation noise model for transform domain Wyner-Ziv video coding. *Journal of Data Acquisition & Processing*, 2009, 24(5): 609-614(in Chinese)
(王凤琴, 樊养余, 赵炯, 刘元魁. 基于变换域 Wyner-Ziv 视频编码的相关噪声模型. *数据采集与处理*, 2009, 24(5): 609-614)
- [16] Maugey T, Gauthier J, Pesquet-Popescu B, Guillemot C. Using an exponential power model for Wyner-Ziv video coding//Proceedings of the ICASSP. Dallas Texas, USA, 2010: 2338-2341
- [17] Brites C, Ascenso J, Pedro J Q, Fernando P. Evaluating a feedback channel based transform domain Wyner-Ziv video codec. *Signal Processing: Image Communication*, 2008, 23(4): 269-297
- [18] Pizurica A, Philips W. Multiscale statistical image models and Bayesian methods//Proceedings of SPIE Conference on Wavelet Applications in Industrial Processing, 2003, 5266: 60-74
- [19] Nadarajah S. Gaussian DCT Coefficient Models. *Acta Applicandae Mathematicae*, 2009, 10(6): 455-472
- [20] Nolan J P. Maximum likelihood estimation and diagnostics for stable distribution//Barndorff-Nielsen O, Mikosch T, Resnick S. *Levy Processes: Theory and Application*. Birkhauser Boston, 2001: 379-400
- [21] Zhang Rong-Yue, Ni Jiang-Qun, Huang Ji-Wu. A robust multi-bits image watermarking algorithm based on HMM in wavelet domain. *Journal of Software*, 2005, 16(7): 1323-1332(in Chinese)
(张荣跃, 倪江群, 黄继武. 基于小波域 HMM 模型的稳健多比特图像水印算法. *软件学报*, 2005, 16(7): 1323-1332)
- [22] Bilmes H A. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Department of EE & CS, U. C. Berkeley; TR-97-021, 1998
- [23] Yin Ming, Zhang Yun, Cheng Liang-Lun, Cai Shu-Ting. Improving on side information estimation for Wyner-Ziv video coding. *Application Research of Computers*, 2009, 26(12): 1001-3695(in Chinese)
(尹明, 章云, 程良伦, 蔡述庭. Wyner-Ziv 视频编码中边信息估计算法改进. *计算机应用研究*, 2009, 26(12): 1001-3695)
- [24] Artigas X, Ascenso J, Dalai M et al. The DISCOVER codec: Architecture, techniques and evaluation//Proceedings of the Picture Coding Symposium. Lisbon, Portugal, 2007
- [25] Varodayan D, Chen D, Flierl M, Girod B. Wyner-Ziv coding of video with unsupervised motion vector learning. *EURASIP Signal Processing: Image Communication*, 2008, 23(5): 369-378



YIN Ming, born in 1975, Ph. D., assistant professor. His research interests include video coding technology, multimedia communication, and embedded image processing system.

Background

This work is partially supported by the National Natural Science Foundation of Guangdong Province under Grant No. 8351009001000002; 211 Project of Guangdong Province under Grant No. 431.

Despite recent advances, the rate-distortion (RD) performance of distributed video coding (DVC) is not yet at the level of traditional motion-compensated predictive coding. The one of critical factors with respect to RD performance is accurately modeling and estimating the correlation channel. The correlation channel is a virtual channel between the side information at the decoder and the original information at the encoder. The side information can be seen the version of the original one contaminated by the correlation noise. It is clearly known that a more accurate description for correlation noise probability distribution indicates fewer parity bits are required to be sent to meet the reconstructed quality. As a result, the RD performance gap between the Wyner-Ziv video codec and the traditional predictive coding solution is to be shortened significantly. However, the virtual correlation channel is non-stationary, which is difficult to model by means of traditional methods.

CAI Shu-Ting, born in 1979, Ph. D., associate professor. His research interest is video coding technology.

XIE Yun, born in 1963, Ph. D., professor. Her research interests include pattern recognition technology, intelligent robot system.

To resolve this problem, other methods corresponding to modeling and adaptive estimating parameter have been proposed in the literature, it still remains a hard challenge and does need to further research on both theory and operation. In this paper, the correlation noise model based on Gaussian mixture distribution is proposed. The statistical features of the correlation noise between WZ frame and corresponding side information are analyzed. It is found that the actual probability distribution does not obey single peak distribution. We then introduce Gaussian mixture distribution to character the noise coefficient histogram. Moreover, the histogram feature is exploited to estimate the model parameter. The experimental results show that the model based on GMM can character the correlation noise more accurately and DVC integrated the proposed model achieves higher RD performance.

The authors have been pursuing research on improving the performance of DVC, such as improvement on side information quality, accurately modeling on correlation noise and integrating the Wyner-Ziv video coding techniques into wireless sensor network, etc.