基于图像块码本模型的监控视频背景参考帧生成方法

张伟，王宇，陈新怡，王延文，景庆阳，雷为民
（东北大学计算机科学与工程学院，辽宁 沈阳 110169）

摘 要：为解决背景参考帧受前景污染严重，以及传输背景参考帧导致的码率突增等问题，针对背景较稳定的监控视频，提出一种以图像块为基本单元的渐进式背景参考帧生成方法。所提方法建立了基于聚类的图像块码本模型，利用基于感知哈希的码元匹配，将视频序列中处于同一位置的图像块进行聚类；利用背景图像区域特性准确检测背景码元；利用码本模型从不同帧中检测出背景图像块生成完整的背景参考帧。实验结果表明，所提方法编码效率相比标准 HM16.20 在亮度分量上提升 17.89%，有效提升了背景参考帧生成质量，且时间复杂度满足视频实时性需求。

关键词：监控视频；背景建模；视频编码；码本模型；背景参考帧
中图分类号：TN92
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2023003

Background reference frame generation method for surveillance video based on image block codebook model

ZHANG Wei, WANG Yu, CHEN Xinyi, WANG Yanwen, JING Qingyang, LEI Weimin
School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China

Abstract: To solve the problems that the background reference frames are seriously contaminated by the foreground, and the bit rate increases suddenly incurred by the one-time transmission of the background frames, a progressive background frame generation method with image block as the basic unit was proposed for surveillance video application. An image block codebook model based on clustering was formulated. The image blocks at the same position in the video sequence were effectively clustered by using perceptual hash-based element matching. The background symbol was accurately detected by using the characteristics of the background image area. A complete background frame was produced by extracting the background blocks in different frames based on the codebook model. Experimental results demonstrate that the proposed method achieves 17.89% coding efficiency for luma component compared with standard HM16.20, and can effectively improve the quality of the produced background reference frame. Besides, the proposed method complexity meets the real-time requirements of video applications.

Keywords: surveillance video, background modeling, video coding, codebook model, background reference frame

0 引言

随着监控设备开始向超高清时代迈进及监控设备部署数量的不断增加，监控视频数据量的增长速度已远大于视频编码效率的提升速度。海量视频数据给视频存储和传输环节带来了极大挑战，进一步提高监控视频压缩率成为一个亟待解决的问题。

目前，视频监控业务一般采用 H.264/AVC（advanced video coding）[1]、H.265/HEVC（high efficiency video coding）[2]等传统编码标准进行编码

收稿日期：2022−09−08；**修回日期**：2022−12−07
基金项目：国家重点研发计划基金资助项目（No.2018YFB1702000）；中央高校基本科研业务费专项资金资助项目（No.N2216010）
Foundation Items: The National Key Research and Development Program of China (No.2018YFB1702000), The Fundamental Research Funds for the Central Universities (No.N2216010)

压缩。然而，这些通用视频编码标准并没有针对监控业务本身的特性提出有效的解决方案，忽略了监控视频中存在的一种特殊冗余信息，即背景冗余。背景冗余是指由于监控视频一般在固定位置拍摄，视频中的背景信息呈现短时间内固定不变、长时间内缓慢变化或规律性变化的特性。由此可见，传统编码标准的编码效率在监控视频业务方面尚有较大提升空间。

针对如何去除监控视频中存在的背景冗余问题，基于背景的监控视频编码方法[3-4]成为近年的研究热点。该类方法通过对背景图像进行建模，并将背景图像设为长期参考帧来提升编码效率，借助视频编码标准中帧间预测过程的参考帧技术，易部署于各类视频编码器，具有广阔的应用前景。虽然此类方法已取得一定研究成果，但仍存在诸多问题。

1) 编码效率的提升依赖于背景参考帧的生成质量，而传统背景建模技术生成的背景参考帧质量较差，普遍存在前景拖影、前景误判为背景等问题。

2) 传统背景建模技术生成的背景参考帧一般需要高质量编码，并一次性传输至解码端完成参考帧列表同步，易发生码率突增从而引起网络拥塞、时延卡顿等问题，影响用户体验质量。

3) 现有基于背景的监控视频编码方法一般以帧为单位更新背景参考帧。出于对背景更新频率与其码率消耗间的权衡，当部分背景区域发生变化时并不会及时更新，使这部分变化的背景区域得不到及时参考利用。

4) 在视频会议场景中，由于人物遮挡等原因无法获得部分区域的真实背景，传统背景建模方法生成的背景参考帧中包含大量无效的参考区域，对此类背景帧编码传输存在码流浪费问题。

针对上述问题，本文提出一种以图像块为基本单元的渐进式背景参考帧生成方法，借鉴像素码本模型思想，构建以图像块为处理单元的码本模型。在视频编码过程中，利用基于感知哈希的码元匹配过程，将视频序列中处于同一位置的图像块进行聚类；分析并利用背景图像区域特性，准确判定码本中的背景码元。利用码本模型从不同视频帧中检测出干净背景图像块，并渐进式生成完整高质量的背景参考帧。背景参考帧的生成、更新及传输机制均以图像块为单位，过程跨越几十至上百帧，可以有效平滑码流，并避免背景参考帧更新不及时和码流浪费问题。本文主要贡献包括以下 3 个方面。

1) 提出一种渐进式背景参考帧生成方法，利用视频序列中的图像块建立并更新图像块码本模型，从中检测出干净背景图像块，通过提取不同帧中检测到的背景图像块渐进式生成完整的背景参考帧。

2) 提出融入渐进式生成背景参考帧机制的视频编码框架，以 H.265/HEVC 编码为例，在此框架下设计了背景参考帧的生成、更新和传输过程。

3) 将所提方法在 PKU-SVD-A 数据集以及 H.265/HEVC 标准测试序列上进行验证，实验结果表明，所提方法比标准 HM16.20 在 PKU-SVD-A 数据集上实现了 17.89%的编码效率增益，有效提升了背景参考帧质量，且复杂度满足视频实时性需求，适用于背景稳定的监控类和会议视频业务。

1 相关研究

1.1 监控视频压缩技术

1) 基于背景的监控视频编码方法

此类方法的编码效率依赖于背景参考帧的生成质量，因此如何选取或生成背景参考帧是这类方法的主要研究内容。文献[5]最早使用长期参考帧（LTR, long-term reference frame）机制来提升监控视频、会议视频的编码性能。文献[6]将关键帧直接用作长期背景参考帧。文献[7]选取 Skip 编码模式数量最多的编码帧作为背景参考帧。这些研究均从已解码帧中选取背景参考帧，由于视频帧中通常含有前景信息，导致解码后选定的背景参考帧的编码质量不高。

在如何生成背景参考帧方面，文献[8]使用解码后的重建帧训练生成背景参考帧，虽然不需要编码传输背景参考帧，但是由于重建帧存在失真，导致生成背景参考帧的失真更严重。更多研究使用原始图像训练生成背景参考帧[9-11]，这类方法一般使用混合高斯背景建模技术等传统背景建模方法生成背景参考帧，但仍普遍存在前景拖影、前景误判为背景等现象，而且一次性传输至解码端易发生码率突增问题。为了解决码率突增问题，文献[12]提出了基于双背景参考帧的编码方案，对原始图像和重建图像分别训练生成 2 个背景帧，编码传输 2 个背景帧之间的残差帧完成编解码器同步，但是背景帧更新仍以帧为单位，小部分背景区域发生变化时不会得到及时更新和参考利用。

有些学者通过建立背景字典库的方式消除背景冗余。对于卫星视频编码，文献[13]以谷歌地球中存储的图像作为背景字典库，根据卫星所处地理位置坐标在背景字典库中搜索背景参考帧。对于交

通监控视频编码，文献[14]提前建立车辆库模型用于后续的预测编码；文献[15]通过视频采集过往车辆和背景信息建立车辆库和背景库，编码时在所建库中检索车辆和背景作为参考进行预测编码。

2) 基于感兴趣区的监控视频编码方法

基于背景的监控视频压缩方法侧重于提高监控视频的压缩率，而基于感兴趣区（ROI, region of interest）的编码方法则是基于人眼视觉特性，更侧重于在有限的码率下合理分配码率实现提高主观感受质量的目的[16]。其中，准确提取感兴趣区是该类方法的研究重点。感兴趣区可以通过手动设置、眼球跟踪设备记录或者内容识别算法预测得出[17-18]。基于深度学习的 ROI 识别是当前较有效的手段。感兴趣区包括中心区域 ROI、人脸 ROI、字幕 ROI 等。鉴于监控视频中能够引起人们视觉注意的是人、车辆等特定对象的变化，大部分研究将运动目标作为监控视频的感兴趣区。文献[19]提出一种运动目标检测方法，根据运动信息将对象与背景分离，不需要训练序列即可同时完成对象检测和背景估计，但是当前景在视频序列中有短暂停留时，该方法会将静止的前景过拟合为背景区域，从而出现漏检现象。

我国自主制定的安防监控音视频编码标准（AVS-S, audio video coding standard for surveillance）提供了基于灵活条带和条带集的 ROI 编码方法[20]，鉴于许多监控应用的监控区域是固定的，压缩标准提供了可以预先设定 ROI 的交互接口。其他视频编码标准没有针对 ROI 特定设计，将视频画面中每个像素看成同等重要，研究表明 ROI 编码对 H.265/HEVC 等标准同样可提升编码性能[21-23]。

3) 基于语义的会议视频编码方法

此类方法认为传统手工设计的混合编码框架下的视频压缩性能已经趋于极限，设想突破传统方法中以视频低层特征（如颜色、运动、纹理）表示视频的局限性，对视频中的高级语义信息进行特征提取，在解码端利用这些高级语义特征重建视频，例如，提取面部关键点[24-25]对人脸及表情变化进行建模，提取骨骼关键点[26]对人体运动信息进行表征，解码端通过生成式对抗网络利用解码后的关键点信息重建面部或者人体姿态。在视频会议场景中，多项研究提取人物边缘信息[27-28]，在解码端利用 pix2pix[29]等重建网络还原人物信息。上述研究在简单场景中可获得一定的压缩增益，但是对非人脸及复杂的前景对象无法获得理想的效果。

1.2　背景建模技术

基于背景的监控视频编码研究中，很多方法利用传统背景建模技术生成背景图像。

1) 混合高斯背景建模算法

单高斯背景建模算法[30]受光线等各种干扰因素影响导致稳健性较差，在此基础上诞生了混合高斯背景建模算法[31-33]。混合高斯背景建模算法中，每个像素点的分布是多个高斯分布模型的叠加状态，如闪烁的屏幕、摇摆的树叶等，在长期观测下像素点会在多个像素值附近聚集，为每个像素点聚集处设置一个高斯分布表示。该算法仍然存在很多问题，例如，停止运动或者移动缓慢的前景在长期观测下也会聚集大量像素点，产生一个以该前景像素点为中心的高斯分布模型，导致前景被误判为背景，生成的背景图像也会出现前景拖影等现象，如图 1 方框所示。另外，在背景建模过程中使用多个高斯分布模型，涉及大量浮点运算，导致时间复杂度较高。

(a) Crossroad序列　　(b) Classover序列

图 1　混合高斯背景建模算法生成的背景图像

2) 中值滤波法

基本的中值滤波法使用当前帧的前 N 帧图像的中值作为背景帧[34]，改进方法使用前 N 帧图像、下采样图像和之前计算出的中值加权得出背景帧[35]，增加背景模型的稳定性。此方法的缺点是需要较大缓存空间，在实际应用中尤其是在微型设备上难以实现。AVS-S 将训练视频帧划分成若干个数据段，使用分段加权滑动平均值法生成背景参考帧。这种方法虽然时间复杂度较低，但是生成的背景图像仍然存在受前景污染严重的问题。

2　基于渐进式生成背景参考帧的视频编码框架

干净的背景参考帧是保证监控视频能够获得理想压缩性能的关键因素。已有的基于背景的监控视频编码方法以图像帧为单位编码、传输和更新背景参考帧，与此不同，本文提出以图像块为单位，从多个视频帧中检测出背景图像块渐进式地合成背景参考帧，

具体过程如下。在视频编码过程中，构建图像块码本模型，利用基于感知哈希的码元匹配过程，将视频序列中处于同一位置的图像块进行聚类；利用背景图像区域特性，准确判定码本中的背景码元。利用码本模型，检测视频帧中与背景码元相匹配的干净背景图像块，并对首次出现的背景图像块进行标记和高质量编码，解码后提取出带标记背景图像块，替换背景参考帧中的对应区域。这种方式以图像块为单位渐进式生成高质量背景参考帧，并且在编码过程中有效避免了一次性传输背景帧导致的码率突增问题。

本节介绍融入渐进式生成背景参考帧机制的视频编码整体框架。以 H.265/HEVC 编码为例，融入渐进式生成背景帧的视频编码框架如图 2 所示，本文以图像块为单元的背景参考帧生成方式与以图像块为编码单元的标准视频编码框架易于结合，涉及生成背景参考帧部分的编码流程如下。

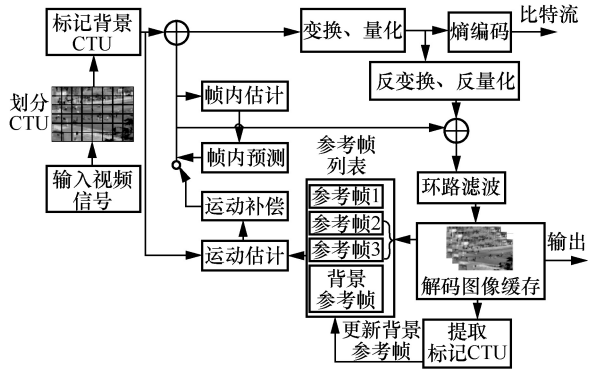


图 2 融入渐进式生成背景帧的视频编码框架

步骤 1 背景参考帧初始化。将 H.265/HEVC 编码器编码的第一个 I 帧初始化为背景参考帧，并将其插入参考帧列表作为长期背景参考帧。

步骤 2 背景图像块检测与标记。将待编码图像帧分割成互不重叠的图像块，分割方式与编码器划分编码树单元（CTU, coding tree unit）方式一致；以图像块为处理单元，利用码元匹配过程更新图像块码本模型；若待编码图像块与码本中的背景码元相匹配，则为背景图像块，为避免对背景参考帧的同一区域重复编码和更新，仅标记首次匹配的背景图像块。

步骤 3 视频帧编码。对带标记背景图像块进行高质量编码，对其他图像块则以默认参数编码。高质量编码可通过减小量化参数来实现，本文第 4 节实验部分带标记背景图像块的量化参数相对于默认量化参数降低 10。

步骤 4 渐进式生成背景参考帧。编解码器重建编码帧并存入解码图像缓冲区后，提取带标记背景图像块，用其替换长期背景参考帧中的对应区域，为后续编码帧提供更高质量的参考帧。

3 基于图像块码本模型的背景帧生成方法

从视频帧中检测到干净、真实的背景图像块是本文渐进式生成背景参考帧方法的核心目标。借鉴像素码本模型[36-37]的聚类思想，本文提出图像块码本模型用于检测视频帧中的背景图像块。像素码本模型是一种以像素为处理单位的前景背景分割算法，一个码本包含若干码元，用于描述一个连续采样的像素。在训练阶段，根据像素值大小放入不同码元，利用背景像素通常比前景像素出现更频繁的特性，提取能够代表背景像素值范围的码元即背景码元；在检测阶段，将与背景码元像素值范围相匹配的像素检测为背景像素。与像素码本模型不同，本文方法不再以像素而是以图像块为基本单位进行统计建模。

3.1 图像块码本模型

图像块码本模型利用码本描述连续采样的图像块，为视频帧的每个图像块位置维护一个码本，码本中包含若干码元，码元为候选背景图像块的集合，通过图像块间的相似度匹配将序列中处于同一位置的图像块进行有效区分并放入不同码元，使其中某个码元所包含的图像块均为真实、干净的背景图像块，称为背景码元。码本分析过程定期从码本中检测并更新背景码元。码本成员信息和码元成员信息如表 1 和表 2 所示。

表 1 码本成员信息

成员	含义
$ce_1, ce_2, \cdots, ce_L$	该码本包含的 L 个码元
L	该码本包含的码元数量
f_{max}	该码本中各码元属性 f 的最大值
λ_{max}	该码本中各码元属性 λ 的最大值
$flag_{bgce}$	该码本是否已建立背景码元，值为 0 或 1

表 2 码元成员信息

成员	含义
mat	中心图像块，为该码元中各图像块的平均图像块
q	该码元最后一次更新图像块的时间
f	该码元包含的图像块数量
λ	该码元中时间相邻图像块之间的间隔帧数累加值
isBgce	该码元是否为背景码元，值为 0 或 1
isMark	该码元是否已标记，值为 0 或 1

图像块码本模型工作过程如下。

步骤 1　码本初始化。为视频帧的每个图像块位置维护一个码本，在视频编码开始时，所有码本初始化为空。

步骤 2　图像块的运动度量。评估待编码图像块的运动情况，若运动度量值小于设定阈值，则判断其为潜在背景块。

步骤 3　码元匹配。将潜在背景块与其所处位置对应码本进行码元匹配，若匹配成功，则将该潜在背景块放入匹配码元并更新码元参数；若匹配失败或码本为空，则新建一个码元并将该潜在背景块放入新建码元。

步骤 4　码本分析。在视频编码过程中，对每个码本周期性地执行该过程，更新码本的背景码元，并清理没有潜力的码元，精简码本模型规模。

在视频编码过程中，利用视频序列中的待编码图像块通过步骤 2 至步骤 4 持续更新图像块码本模型。在步骤 3，若与潜在背景块匹配成功的是背景码元，则该潜在背景块被判定为背景图像块。为了防止背景参考帧的重复编码和更新，编码器仅对首次匹配的背景图像块进行一次高质量编码。

3.2　图像块的运动度量

运动特性是区分图像前景和背景的重要手段。固定位置拍摄的监控视频中，背景区域较稳定，而运动目标所在区域在相邻帧中像素值存在明显差异。为了提高背景图像块检测精度，本文利用运动特征对图像块进行预处理，粗略筛选出潜在背景块，由于此阶段对精度要求不高，选用复杂度较低的帧间差分法来评估图像块的运动情况。图像块的运动度量如图 3 所示。

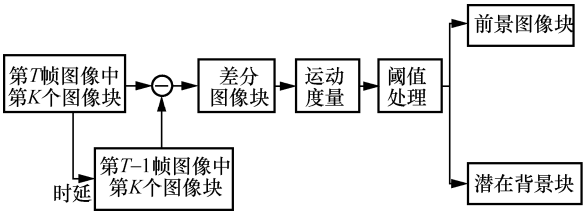


图 3　图像块的运动度量

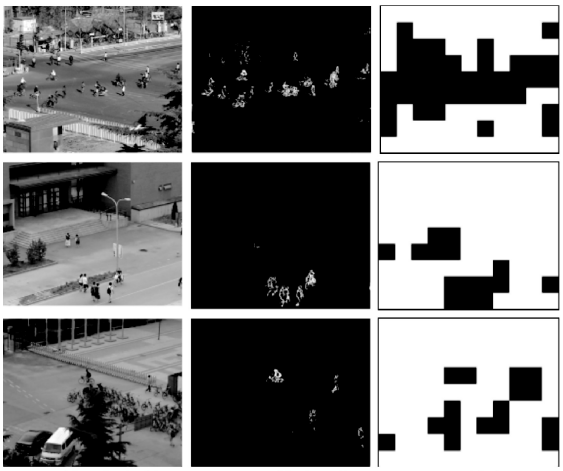
对相邻帧中同一位置图像块做差分运算得到差分图像块 D_T^K，将 D_T^K 中的像素均值作为图像块的运动度量值 V_T^K，利用 V_T^K 与设定阈值的大小关系将图像块划分为前景图像块与潜在背景块。

$$D_T^K(i,j)=F_T^K(i,j)-F_{T-1}^K(i,j) \tag{1}$$

$$V_T^K=\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|D_T^K(i,j)\right|}{N^2} \tag{2}$$

其中，$F_T^K(i,j)$ 和 $F_{T-1}^K(i,j)$ 分别表示视频帧 F_T 和 F_{T-1} 中第 K 个图像块中坐标为 (i,j) 的像素值，N 表示图像块的边长。

视频帧的运动度量效果示意如图 4 所示，图 4(c) 中白色区域为潜在背景块，黑色块为前景块。



(a) 输入视频帧　(b) 差分图像　(c) 输出

图 4　视频帧的运动度量效果示意

由于摄像头的抖动及噪声干扰，导致画面的像素值变化整体偏高，区分噪声和真实运动较困难，不宜将运动阈值设置过低。另一方面，缓慢运动或突然停止运动的前景均呈现较低的运动度量值。为此，仅将图像块的运动度量作为预处理，后续的码元匹配与码本分析过程精确检测背景图像块。

3.3　基于感知哈希的码元匹配

码元匹配过程将视频序列中连续采样的图像块分配到相应码本的不同码元，由于噪声干扰和摄像头抖动等诸多因素，不能简单地根据图像块间差值的大小来实现码元匹配。为了提高稳健性，本文提出基于感知哈希的码元匹配算法。与传统加密哈希不同，相似输入的感知哈希值也是相似的。利用感知哈希值的可比较性，将图像块感知哈希值之间的汉明距离作为图像块间的相似度度量，并据此将同一位置的图像块分配至相应码本的不同码元。图像块 X 的感知哈希值计算如下

$$H(X)=\mathrm{VEC}(\mathrm{BIN}(\mathrm{AVG}(\mathrm{DCT}_{8\times 8}(\mathrm{SCALE}(\mathrm{GRAY}(X))))))\tag{3}$$

其中，GRAY 表示对图像块进行灰度化处理；SCALE 表示对图像块进行缩放处理；$\mathrm{DCT}_{8\times 8}$ 表示计算图像块的 DCT 系数矩阵，并选取系数矩阵左上角的 8×8 矩阵，保留 DCT 系数矩阵的低频信息；AVG 表示计算矩阵元素的平均值；BIN 表示按照阈值完成系数矩阵的二值化；VEC 表示将二值化的矩阵按照固定顺序拉伸为向量形式的哈希值。

起始阶段所有码本初始化为空。设待处理的潜在背景块为第 t 帧第 i 个图像块 x_t^i，图像帧第 i 个图像块位置对应码本 CB_i。码元匹配过程简单描述如下。计算 x_t^i 感知哈希值 $H\left(x_t^i\right)$ 与码本 CB_i 中各码元 ce_j $(1\leqslant j\leqslant L)$ 感知哈希值 $H\left(\mathrm{ce}_j.\mathrm{mat}\right)$ 之间的汉明距离，选取具有最小汉明距离的码元 ce_k，若最小汉明距离大于阈值 D 则表明匹配失败，需要新建一个码元并将 x_t^i 放入新建码元；否则表明匹配成功，将 x_t^i 放入该码元 ce_k，更新码元 ce_k 的成员信息，其中，中心图像块 mat 更新为

$$\mathrm{mat}=\frac{f\mathrm{mat}+x_t^i}{f+1}\tag{4}$$

3.4　码本分析

码本分析过程周期性地检测或更新码本中的背景码元，并精简码本模型规模。工作原理是依据本文分析观察到的背景图像块的 3 个特征。

特征 1　背景图像块出现频率 f 高。图像区域划分示意如图 5 所示，按照前景出现频率可将图像帧划分为 2 类区域，I 类区域中前景出现频率低，多为干净背景，而 II 类区域中前景出现频率较高。为此，I 类区域内的码本中，背景码元与其他码元相比包含更多的图像块，且图像块出现频率 f 较高。图像序列 I 类区域中的背景码元如图 6 所示，视频序列中的右上角图像块始终为背景图像块，该位置码本中背景码元的图像块出现频率 f 将近似于视频帧数。由此 I 类区域内的码本可以利用图像块出现频率 f 区分出背景码元。

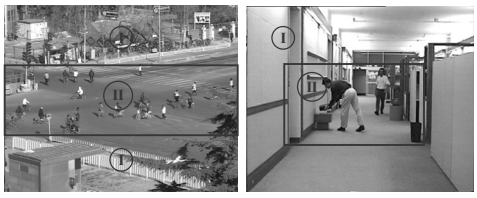


图 5　图像区域划分示意

图 6　图像序列 I 类区域中的背景码元

特征 2　背景图像块间隔式出现。图像序列 II 类区域中的背景码元如图 7 所示。II 类区域中前景出现频繁，背景码元和其他码元的图像块出现频率无显著差异。有些场景中，背景图像块呈现出间隔式出现的规律，如图 7 所示的走廊区域的背景图像块。本文将这种规律量化为码元中时间相邻图像块之间的间隔帧数累加值 λ，II 类区域内的码本可借助 λ 区分出背景码元。

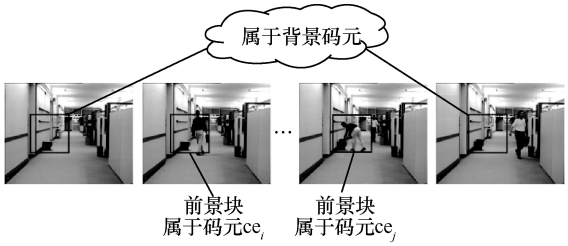


图 7　图像序列 II 类区域中的背景码元

特征 3　背景图像块的纹理复杂度较低。在一些复杂场景，如十字路口等红灯的汽车、突然停下的行人等，根据上述 2 种特性区分背景码元将发生误判。观察可知，背景图像与前景图像相比内部一般是平坦的。本文使用灰度共生矩阵方法[38]统计图像的纹理特征信息，选用灰度共生矩阵的熵值（ENT, entropy）和逆差矩（IDM, inverse different moment）特征表示图像块的纹理复杂度。其中，ENT 表示图像块内容的随机性，反映了图像块的信息量和复杂度，ENT 越大表明图像纹理越复杂；IDM 反映图像块分布平滑度的度量，IDM 越大表明图像越均匀。

$$\mathrm{ENT}=-\sum_i\sum_j p(i,j)\lg p(i,j)\tag{5}$$

$$\mathrm{IDM}=\sum_i\sum_j\frac{1}{1+(i-j)^2}p(i,j)\tag{6}$$

其中，$p(i,j)$ 表示特定位置关系下的像素对的频率，i 和 j 为 2 个像素点的灰度量化值。

利用第 4 节使用的监控视频公开数据集验证特征 3，图像块的纹理复杂度比较如图 8 所示，受前景污染的图像块的 ENT 均大于干净的背景块，

IDM 均小于干净的背景块。本文通过图像块的 ENT 和 IDM 区分干净背景块与受前景污染的图像块，降低前景误判为背景的概率，并在误判发生后及时更正背景帧中的错误背景块。

根据上述特性，利用码本中各码元包含的图像块数量、相邻图像块之间的间隔帧数累加值以及图像块纹理复杂度等指标区分出背景码元。具体为根据特征 1 和特征 2，选取码本中包含图像块数量最多且其值大于阈值 Th_f 的码元，以及选取码本中相邻图像块之间间隔帧数累加值最大且其值大于阈值 Th_x 的码元，将其判定为潜在背景码元；若码本尚无背景码元，且只有一个潜在背景码元，将潜在背景码元直接判定为背景码元，否则根据特征 3，将图像块纹理复杂度最低的潜在背景码元判定为背景码元；若码本已有背景码元，则从潜在背景码元和原有背景码元中选取纹理复杂度最低的图像块，将其设定为新的背景码元。

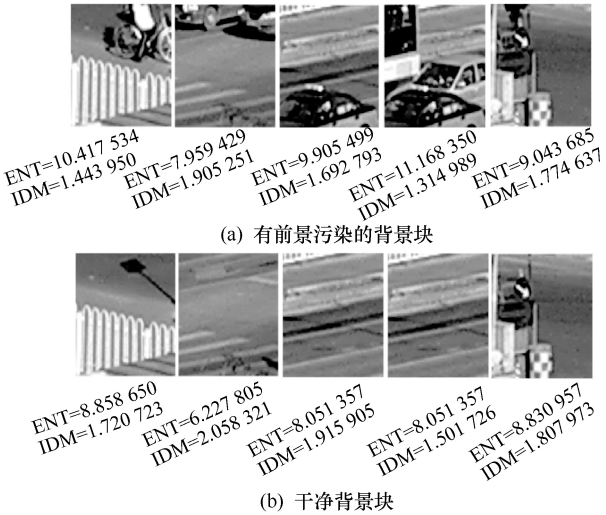


图 8　图像块的纹理复杂度比较

考虑到模型的空间和时间复杂度，在码本分析的最后一步精简码本模型的规模，只保留 f 和 λ 值较大的码元。本文第 4 节实验对于已有背景码元的码本，只保留背景码元以及 f 和 λ 值最大的前 2 个码元；对于尚无背景码元的码本，保留码本中 f 和 λ 值最大的前 5 个码元。

4　实验与结果分析

为了验证算法有效性，对背景参考帧的生成质量以及将其分别应用于监控视频和会议视频 2 种场景中的编码性能进行分析。

4.1　背景参考帧生成质量

选取北京大学监控视频公开数据集 PKU-SVD-A[39]中的若干序列作为测试数据，与基于像素码本模型的背景建模方法（CB）[36]，以及最常用的 2 种经典背景建模方法——高斯混合模型（GMM, Gaussian mixture model）和分段加权滑动平均值（SWRA, subsection weighted moving average）法进行对比。3 个测试序列如图 9 所示，场景分别为道路十字路口、教室门口和校园。

(a) Crossroad序列　(b) Classover序列　(c) Campus序列

图 9　测试序列示意

图 10～图 12 为 3 个测试序列应用不同方法在第 100 帧、200 帧、300 帧时生成的背景参考帧。由于本文方法渐进式生成背景参考帧，在测试过程中仍未生成背景的区域以黑色图像块表示。最右列为生成背景参考帧的部分区域的细节展示。

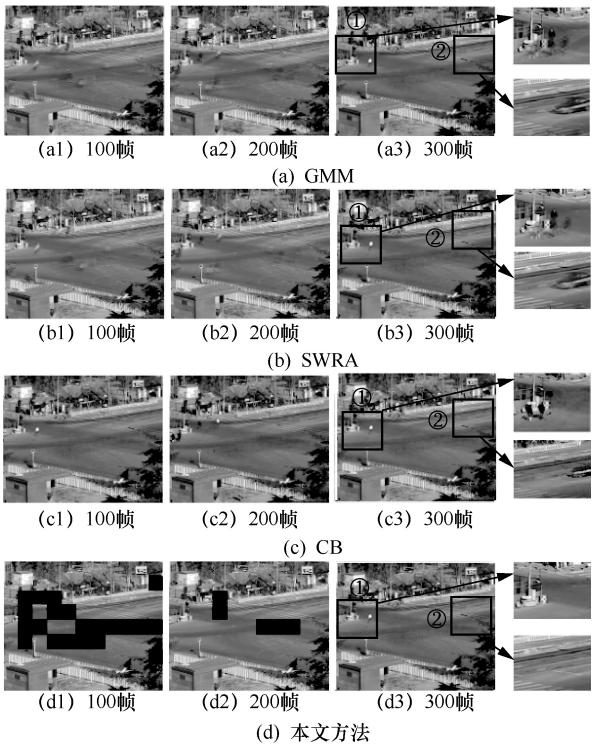


图 10　Crossroad 序列的背景参考帧

从图 10～图 12 中方框圈出的区域①可看出，GMM、SWRA 和 CB 生成的背景参考帧含有大量人物拖影，即存在前景拖影现象。从区域②可看出，这

3 种方法同时还存在前景误判为背景的现象，图 10 中等待红灯的出租车以及图 11 中停留交谈的 2 个行人均被误判为背景。这些误判导致生成的背景参考帧的真实性下降，严重影响其作为背景参考帧带来的编码性能增益。根本原因在于 GMM、SWRA 和 CB 均属于以像素为处理单位的建模方法。

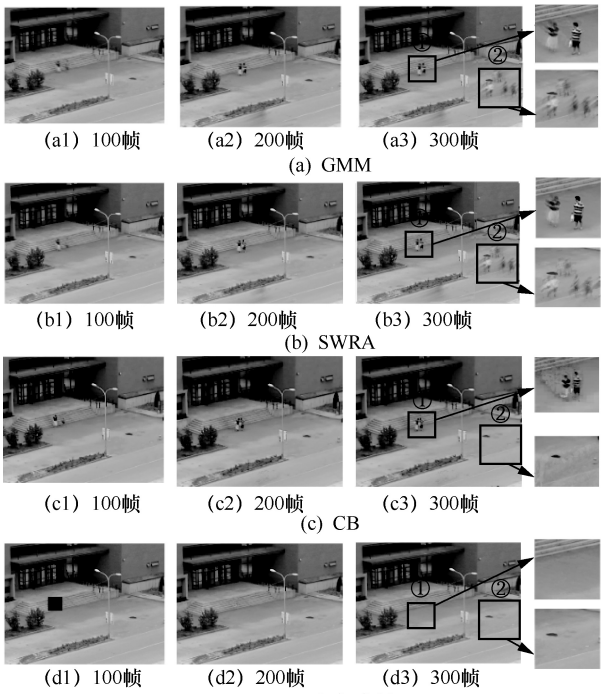


图 11　Classover 序列的背景参考帧

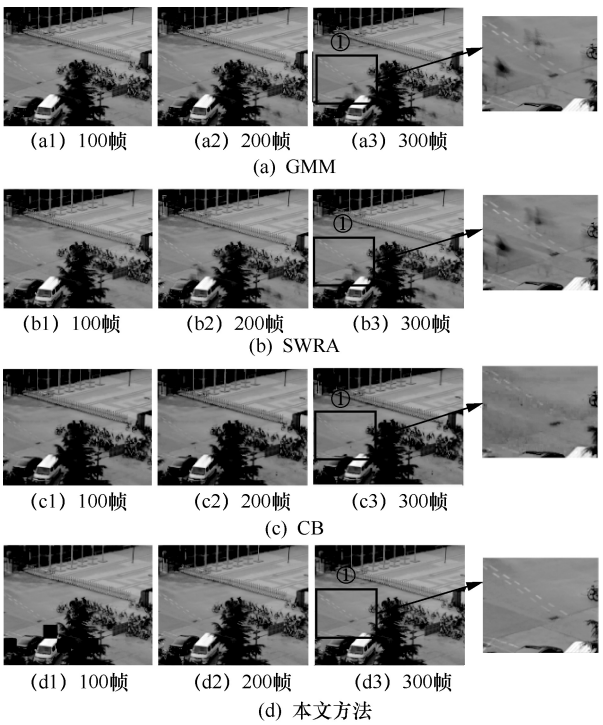


图 12　Campus 序列的背景参考帧

从代表仍未生成背景的黑色图像块可以看出，相对 I 类区域，本文方法对 II 类区域（即有前景干扰区域）较晚提取出背景图像块，但在第 300 帧之前渐进式提取出所有真实背景图像块，生成了完整的、干净真实的背景参考帧。本文方法不存在前景拖影和前景误判为背景的问题，原因在于本文方法不是逐像素按照某种规则生成背景帧，而是从多个视频帧中以图像块为单位提取出背景块。

选择型号为 AMD Ryzen 5 3500U 的 CPU 平台对不同方法的时间复杂度进行比较，结果如表 3 所示。从表 3 可以看出，本文方法平均时间复杂度最低，运行包含 500 帧图像的测试序列平均仅需 16.061 s，即每秒平均处理 31.13 帧，满足视频编码的实时性处理需求。

表 3　　不同方法时间复杂度对比

测试序列	运行 500 帧所需时间/s			
	GMM	SWRA	CB	本文方法
Crossroad	20.466	17.321	19.313	15.463
Campus	20.175	17.512	19.842	16.385
Classover	20.023	17.608	19.815	16.337
平均	**20.221**	**17.480**	**19.657**	**16.061**

会议视频序列的生成背景参考帧如图 13 所示。在会议视频中，由于人物遮挡导致真实的背景区域始终无法获得，GMM、SWRA 和 CB 生成的背景参考帧会包含大量无效区域。本文方法以图像块为单元渐进式合成及更新背景参考帧，可有效避免编码和传输无效的背景区域。从图 13 可以看出，本文方法生成的背景参考帧不包括被人物遮挡的区域。

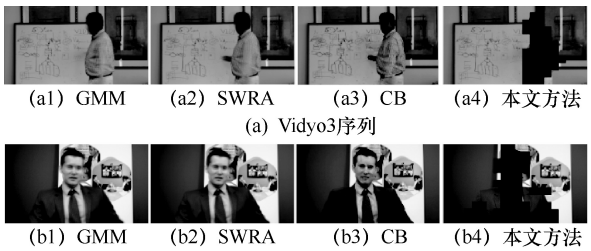


图 13　会议视频序列的生成背景参考帧

4.2　监控视频编码性能分析

为验证在监控视频中的编码性能，将算法集成到 H.265/HEVC 的标准测试软件 HM16.20，并选用满足实时性要求的低时延配置，以 22、27、32 和 37 这 4 个量化参数（QP）对测试视频进行编码，以广泛使用的 BD-rate 作为性能评价指标，与标准

HM16.20 进行对比分析。

测试视频全部选自 PKU-SVD-A 数据集，同时该测试视频也是 AVS 监控档次的测试序列，包括序列 Crossroad、Classover、Campus、Bank、Overbridge、Office 等，均为固定机位拍摄，涵盖校园、马路、办公室等多种生活常见场景，监控视频测试序列如图 14 所示。视频序列帧率为 30 fame/s，帧数为 1 500 帧，采样格式为 4:2:0。为便于算法实现，将视频长宽裁剪成 64 的倍数，即分辨率由 720×576 裁剪成 704×576。

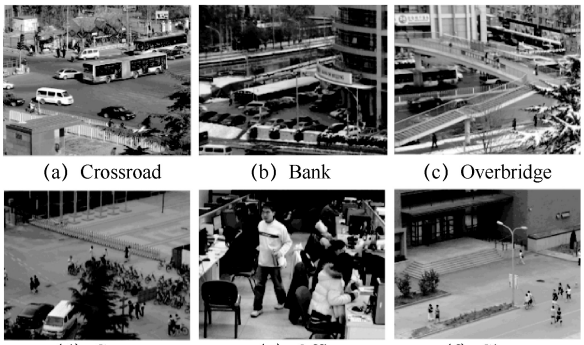


图 14　监控视频测试序列

监控视频测试序列上的 BD-rate 增益如表 4 所示，表 4 中数据为负表示增益。在评价编码性能时一般以亮度分量 *Y* 的增益为主，其他分量的增益作为参考。在分量 *Y* 上，所有测试序列均获得正向增益，平均编码增益为17.89%，即相同视频编码质量下，比HM16.20平均节省 17.89%的码流，最大增益为 Overbridge 序列的 29.54%，最小增益为 8.37%，来自前景占比较大及运动较为频繁的 Crossroad 序列。在分量 *U*、*V* 增益和 YUV 整体增益上，也取得了良好效果，分别为 76.43%、76.22%和 26.68%的平均增益。在时间复杂度方面，以编码所需时间为衡量标准，对比标准 HM16.20，背景帧生成部分所占时间平均仅增加了 0.98%，可满足编码的实时性要求。

表 4　监控视频测试序列上的 BD-rate 增益

视频序列	BD-rate(本文方法与 HM16.20 相比)			
	Y	*U*	*V*	YUV
Bank	–21.24%	–78.46%	–80.24%	–35.08%
Campus	–20.99%	–75.37%	–79.45%	–26.47%
Classover	–18.02%	–75.20%	–78.32%	–23.84%
Crossroad	–8.37%	–75.60%	–69.32%	–18.61%
Office	–9.19%	–73.79%	–70.50%	–16.10%
Overbridge	–29.54%	–80.18%	–79.46%	–39.95%
平均	–17.89%	–76.43%	–76.22%	–26.68%

2 种方法在不同序列前 150 帧的视频帧比特消耗如图 15 所示。第一帧作为初始背景参考帧需要高质量编码，帧比特消耗为一般 I 帧的数倍、P 帧的几十倍。为清楚地展示 2 种方法在后续帧比特消耗上的差异，图 15 中未给出第一帧图像的大比特消耗。本文方法以图像块为单位渐进式生成背景参考帧，并对检测出来的新背景图像块进行高质量编码，本节实验采用相对于默认量化参数减 10 的量化参数来实现，虽然很多帧的比特消耗略高于 HM16.20 标准软件，但整体码率相近，不存在传统背景建模方法所导致的码率突增问题。

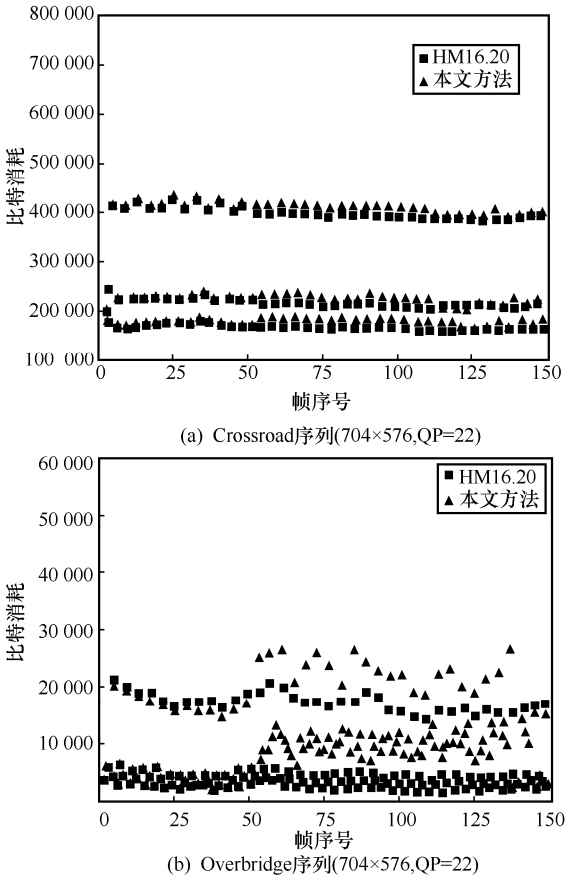


图 15　监控视频帧的比特消耗

监控视频测试序列的率失真曲线如图 16 所示，即码率随分量 *Y* 峰值信噪比（PSNR, peak signal to noise ratio）的变化。从图 16 可以看出，本文方法带来的编码增益更多集中在低码率阶段，例如，Overbridge 序列码率为 130 kbit/s 时，比标准 HM16.20 的 PSNR 值高约 3.5 dB。随着码率升高，帧间参考的时域依赖性降低，背景参考帧的参考价值逐渐下降，导致 2 种方法的编码性能差距逐渐缩小甚至持平。

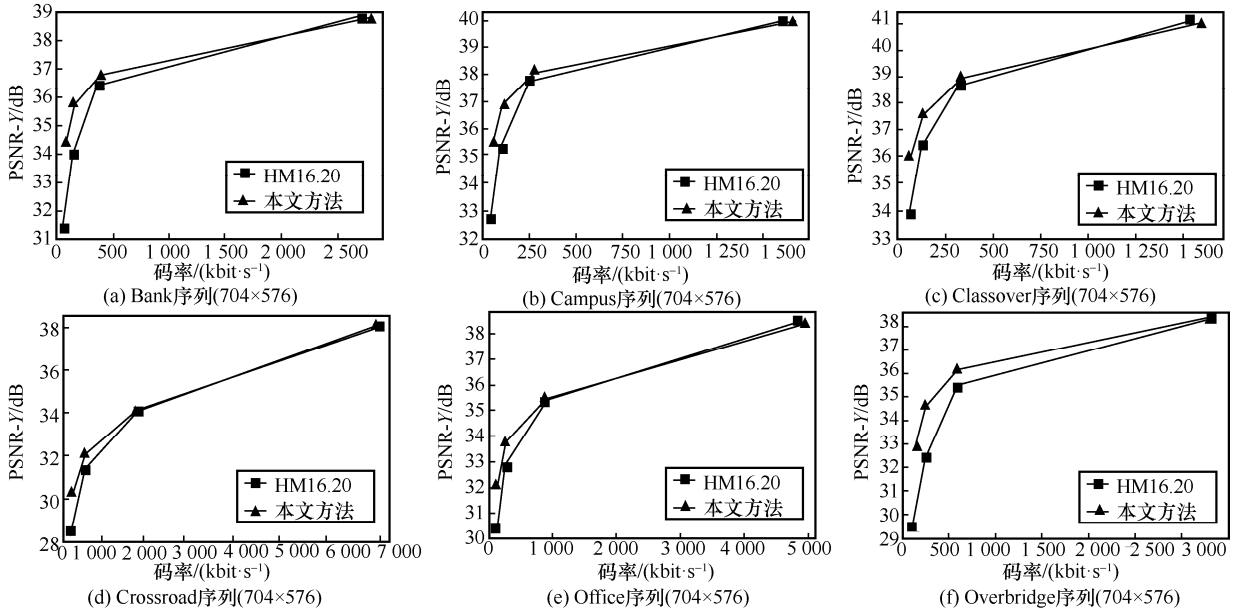
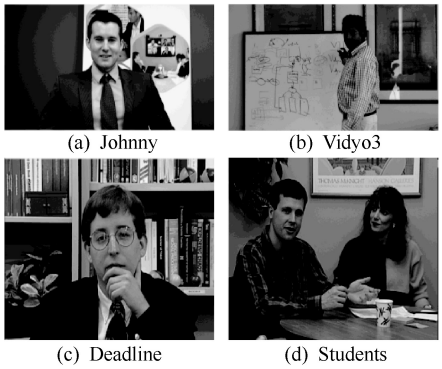


图 16　监控视频测试序列的率失真曲线

4.3　会议视频编码性能分析

会议视频与监控视频类似，一般均为摄像机在固定位置拍摄，具有稳定背景，不同在于前景一般占比较大且多为人像。为验证在会议视频中的编码性能，编码参数设置与监控视频相同，采用相同的低时延配置和 QP，评价基准仍然是标准 HM16.20。会议类测试视频选自 H.265/HEVC 标准测试序列，包括 Deadline 和 Students 这 2 个通用影像传输格式（CIF）序列，以及 Johnny 和 Vidyo3 这 2 个高清晰度（HD）序列，如图 17 所示，视频测试序列相关信息如表 5 所示。



(a) Johnny　(b) Vidyo3　(c) Deadline　(d) Students

图 17　会议视频测试序列

表 5　会议视频测试序列的相关信息

序列名	分辨率	帧率/(fame·s^{-1})	帧数/次	采样格式
Deadline	320×256	30	1 000	4:2:0
Students	320×256	30	1 000	4:2:0
Johnny	1 280×704	60	600	4:2:0
Vidyo3	1 280×704	60	600	4:2:0

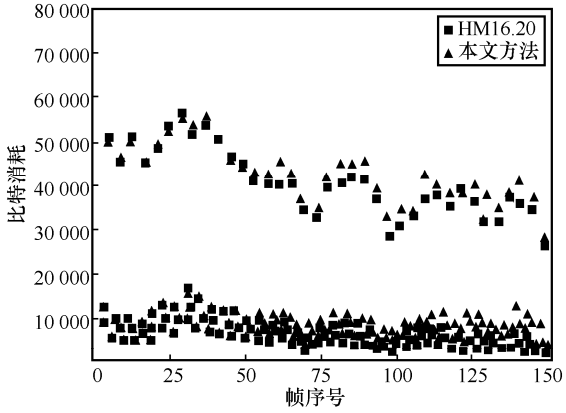
会议视频测试序列上的 BD-rate 增益如表 6 所示。在分量 Y 上，所有测试序列均获得正向增益，平均编码增益为 17.59%，验证本文方法对会议视频同样有效。具体来看，最大增益为 CIF 序列 Students 的 29.29%，最小增益为 HD 序列 Johnny 仅获得 8.57%，该序列背景过于简单，即使在没有背景参考帧参与的情况下该区域的编码失真也较小，所以融入高质量背景参考帧后编码性能提升并不显著，而其他序列的背景区域均相对更复杂，编码性能提升较显著。在分量 U 和 V 上，分别取得 42.15% 和 43.48% 的更大增益效果，YUV 的整体增益为 19.92%，因此在多个测量分量上均达到稳定有效的编码增益。在时间复杂度方面，相比标准 HM16.20，背景帧生成部分所占时间平均仅增加了 1%，可满足编码的实时性要求。

表 6　会议视频测试序列上的 BD-rate 增益

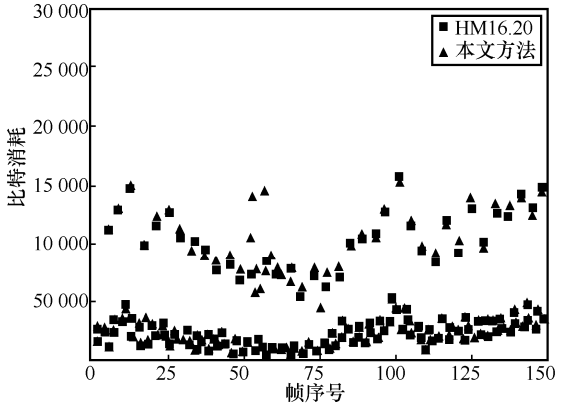
视频序列	BD-rate(本文方法与 HM16.20 相比)			
	Y	U	V	YUV
Deadline	−15.91%	−34.03%	−32.79%	−17.69%
Students	−29.29%	−48.59%	−48.35%	−31.08%
Johnny	−8.57%	−44.06%	−33.28%	−11.64%
Vidyo3	−16.61%	−41.90%	−59.51%	−19.27%
平均	−17.59%	−42.15%	−43.48%	−19.92%

会议视频帧的比特消耗如图 18 所示。2 种方法整体编码情况相似，不存在大的码率突变点，虽然在 Deadline 序列中的第 50 帧到 60 帧之间本文方法较

HM16.20 标准略有增高，但增幅不大，进一步验证本文方法可有效避免传统方法导致的码率突增问题。



(a) Vidyo3序列(1280×704,QP=22)



(b) Deadline序列(320×256,QP=22)

图 18　会议视频帧的比特消耗

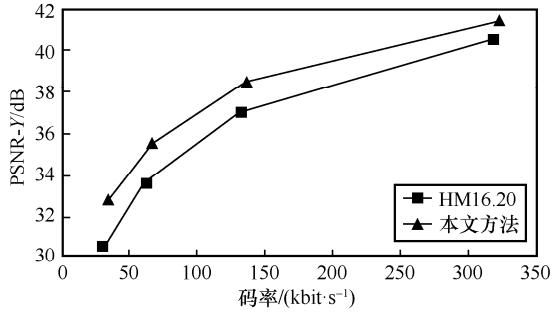
会议视频测试序列的率失真曲线如图 19 所示。大部分测试序列在整个码率阶段均明显高于标准 HM16.20，其中，2 个 CIF 序列的性能增益在整个码率阶段表现更加均匀，而 2 个 HD 序列则与监控视频的率失真曲线类似，编码增益随着码率的降低而增大，说明在低带宽环境下，编码性能的提升将能带来更显著的用户体验质量提升。

5　结束语

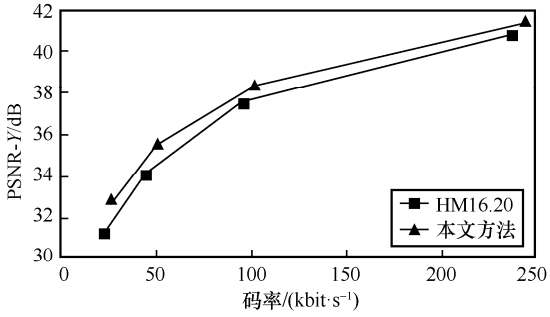
针对传统方法生成的背景参考帧质量差以及传输过程中码率突增等问题，本文提出一种渐进式背景参考帧生成方法，建立以图像块为基本单元的码本模型，利用帧间差分法评估图像块的运动情况；利用感知哈希对同一位置图像块进行聚类分析；利用码元中图像块数量、相邻图像块之间的间隔帧数累加值以及图像块纹理复杂度等指标检测背景码元。在编码过程中利用码本模型从不同帧中检测出干净、真实背景块来生成及更新高质量背景参考帧。结果实验表明，本文方法可以生成高质量背景参考帧，时间复杂度可以满足实时视频需求，可用于背景较稳定的监控类和会话类视频应用。

参考文献：

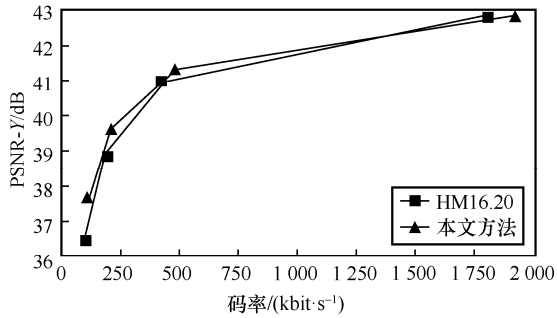
[1] WIEGAND T, SULLIVAN G J, BJONTEGAARD G, et al. Overview of the H.264/AVC video coding standard[J]. IEEE Transactions on



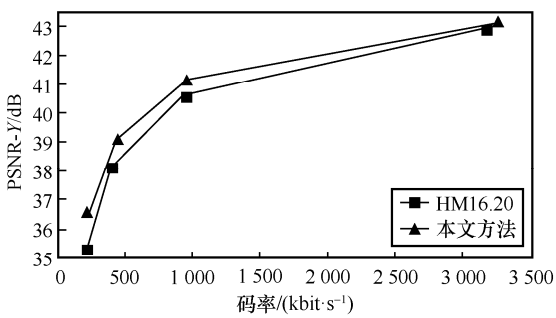
(a) Deadline序列(320×256)



(b) Students序列(320×256)



(c) Johnny序列(1 280×704)



(d) Vidyo3序列(1 280×704)

图 19　会议视频测试序列的率失真曲线

Circuits and Systems for Video Technology, 2003, 13(7): 560-576.

[2] SULLIVAN G J, OHM J R, HAN W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.

[3] ZHANG X G, HUANG T J, TIAN Y H, et al. Background-modeling-based adaptive prediction for surveillance video coding[J]. IEEE Transactions on Image Processing, 2014, 23(2): 769-784.

[4] ZHANG X G, TIAN Y H, HUANG T J, et al. Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling[J]. IEEE Transactions on Image Processing, 2014, 23(10): 4511-4526.

[5] WIEGAND T, ZHANG X Z, GIROD B. Long-term memory motion-compensated prediction[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(1): 70-84.

[6] TUNG C C, YU W H, CHUAN Y T, et al. Single reference frame multiple current macroblocks scheme for multi-frame motion estimation in H.264/AVC[C]//2005 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway: IEEE Press, 2005: 1790-1793.

[7] GORUR P, AMRUTUR B. Skip decision and reference frame selection for low-complexity H.264/AVC surveillance video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(7): 1156-1169.

[8] PAUL M, LIN W S, LAU C T, et al. Video coding using the most common frame in scene[C]//2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Piscataway: IEEE Press, 2010: 734-737.

[9] ZHAO L, WANG S Q, WANG S S, et al. Enhanced surveillance video compression with dual reference frames generation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1592-1606.

[10] CHEN F D, LI H Q, LI L, et al. Block-composed background reference for high efficiency video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(12): 2639-2651.

[11] ZHANG X, HUANG T, TIAN Y, et al. Fast and efficient transcoding based on low-complexity background modeling and adaptive block classification[J]. IEEE Transactions on Multimedia, 2013, 15(8): 1769-1785.

[12] LI H R, DING W P, SHI Y H, et al. A double background based coding scheme for surveillance videos[C]//2018 Data Compression Conference (DCC). Piscataway: IEEE Press, 2018: 420-420.

[13] WANG X, HU R, WANG Z, et al. Virtual background reference frame based satellite video coding[J]. IEEE Signal Processing Letters, 2018, 25(10): 1445-1449.

[14] MA C Y, LIU D, PENG X L, et al. Surveillance video coding with vehicle library[C]//2017 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2017: 270-274.

[15] MA C Y, LIU D, PENG X L, et al. Traffic surveillance video coding with libraries of vehicles and background[J]. Journal of Visual Communication and Image Representation, 2019, 60: 426-440.

[16] NACCARI M, PEREIRA F. Advanced H.264/AVC-based perceptual video coding: architecture, tools, and assessment[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(6): 766-782.

[17] XU J, GUO J, BAO J. A ROI encryption scheme for H.264 video based on moving object detection[C]//2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA). Piscataway: IEEE Press, 2013: 494-497.

[18] LEUVEN S V, SCHEVENSTEEN K V, DAMS T, et al. An implementation of multiple region-of-interest models in H.264/AVC[J]. Signal Processing for Image Enhancement and Multimedia Processing, 2008, 31: 215-225.

[19] ZHOU X, YANG C, YU W. Moving object detection by detecting contiguous outliers in the low-rank representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 597-610.

[20] 马思伟. AVS 视频编码标准技术回顾及最新进展[J]. 计算机研究与发展, 2015, 52(1): 27-37.
MA S W. History and recent development of AVS video coding standards[J]. Journal of Computer Research and Development, 2015, 52(1): 27-37.

[21] MEDDEB M, CAGNAZZO M, PESQUET B P. ROI-based rate control using tiles for an HEVC encoded video stream over a lossy network[C]//2015 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2015: 1389-1393.

[22] ZHANG Z, JING T, HAN J, et al. A new rate control scheme for video coding based on region of interest[J]. IEEE Access, 2017, 5: 13677-13688.

[23] PATEL Z, RAO K R. Image segmentation approach for realizing zoomable streaming HEVC video[C]//2015 International Conference on Science and Technology (TICST). Piscataway: IEEE Press, 2015: 76-82.

[24] OQUAB M, STOCK P, GAFNI O, et al. Low bandwidth video-chat compression using deep generative models[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2021: 2388-2397.

[25] FENG D, HUANG Y, ZHANG Y, et al. A generative compression framework for low bandwidth video conference[C]//2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Piscataway: IEEE Press, 2021: 1-6.

[26] WU Y, HE T, CHEN Z. Memorize, then recall: a generative framework for low bit-rate surveillance video compression[C]//2020 IEEE International Symposium on Circuits and Systems. Piscataway: IEEE Press, 2020: 1-5.

[27] KIM S, PARK J S, BAMPIS C G, et al. Adversarial video compression guided by soft edge detection[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2020: 2193-2197.

[28] HU Y, YANG S, YANG W, et al. Towards coding for human and machine vision: a scalable image coding approach[C]//2020 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2020: 1-6.

[29] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 1125-1134.

[30] BENEZETH Y, JODOIN P M, EMILE B, et al. Review and evaluation of commonly-implemented background subtraction algorithms[C]// 2008 19th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2008: 1-4.

[31] SOBRAL A, VACAVANT A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos[J]. Computer Vision and Image Understanding, 2014, 122: 4-21.

[32] BOUWMANS T, EL B F, VACHON B. Background modeling using mixture of Gaussians for foreground detection-a survey[J]. Recent Patents on Computer Science, 2008, 1(3): 219-237.

[33] SARANLI A. A Gaussian-mixture based approach to spatial image background modeling and compensation[C]//2007 15th European Signal Processing Conference (EUSIPCO). Piscataway: IEEE Press, 2007: 1457-1461.

[34] LO B P L, VELASTIN S A. Automatic congestion detection system for underground platforms[C]//2001 International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP). Piscataway: IEEE Press, 2001: 158-161.

[35] CUCCHIARA R, GRANA C, PICCARDI M, et al. Detecting moving objects, ghosts, and shadows in video streams[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1337-1342.

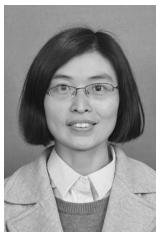
[36] KIM K, CHALIDABHONGSE T H, HARWOOD D, et al. Real-time foreground-background segmentation using codebook model[J]. Real-Time Imaging, 2005, 11(3): 172-185.

[37] DOSHI A, TRIVEDI M. “Hybrid cone-cylinder” codebook model for foreground detection with shadow and highlight suppression[C]//2006 IEEE International Conference on Video and Signal Based Surveillance. Piscataway: IEEE Press, 2006: 19-19.

[38] HARALICK R M, SHANMUGAM K, DINSTEIN I. Textural features for image classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1973, SMC-3(6): 610-621.

[39] GAO W, TIAN Y, HUANG T, et al. The IEEE 1857 standard: empowering smart video surveillance systems[J]. IEEE Intelligent Systems, 2013, 29(5): 30-39.

[作者简介]



张伟（1980– ），女，山东济宁人，博士，东北大学讲师，主要研究方向为多媒体智能信号处理和网络多径传输优化。

王宇（1997– ），男，黑龙江齐齐哈尔人，东北大学硕士生，主要研究方向为多媒体智能信号处理。

陈新怡（1994– ），女，河北承德人，东北大学博士生，主要研究方向为计算机视觉、视频图像压缩编码。

王延文（1998– ），女，辽宁辽阳人，东北大学博士生，主要研究方向为计算机视觉、视频图像压缩编码。

景庆阳（1994– ），女，辽宁沈阳人，东北大学博士生，主要研究方向为计算机视觉、视频图像压缩编码。

雷为民（1969– ），男，山西平遥人，博士，东北大学教授，主要研究方向为多媒体智能信号处理、网络多径传输优化和工业实时通信技术等。