Temporal Sampling Based Multiple Description Video Coding for Scenes Switching

Mengmeng Zhang*, *Member IEEE*, Huihui Bai+

* *North China University of Technology*
No.5 Jinyuanzhuang Road, Shijingshan
Beijing, 100144, China
zmm@ncut.edu.cn

+ *Beijing Jiaotong University*
No.3 Shang Yuan Cun,Hai Dian
Beijing, 100044, China
luckybhh@gmail.com

Due to network congestion and delay sensibility, it is always a great challenge for video transmission over lossy network. Multiple description coding (MDC) is an attractive approach to solve this problem. It can efficiently combat packet loss without any retransmission thus satisfying the demand of real time services and relieving the network congestion. In view of perfect compatibility with the standard source and channel codec, temporal sampling based MDC has become a better choice for practical applications. However, for the frames switching from one scene to another temporal correlation may be destroyed by sampled in temporal domain, which may result in the false estimation when the related frames are lost at the side decoder. To address this problem, in this paper an improved MD coding based on temporal sampling is proposed to make sure the decoder can work correctly when scenes changing.

It is noted that for the video comprised by different scenes, there is no correlation between the last frame in previous scene and the first frame in subsequent scene even if the two frames are neighboring. Therefore we must find out where scenes switching occurs in the video firstly. For simplicity, the sum absolution difference (SAD) between the pixels of two neighboring frames is utilized as the criterion. Let $p_k(x,y)$ denote the pixel at the coordinates (x,y) in frame F_k. For the frames with $M \times N$ pixels, if $\sum_{x=1}^{M}\sum_{y=1}^{N}\left|p_k(x,y)-p_{k-1}(x,y)\right| \geq T_3$, scenes switching happens between frame F_{k-1} and F_k. When scenes switching appears, the first and the last frame of the current scene will be transmitted over all channels to ensure the intact rebuilding of frames. Then instead of the odd/even frame splitting directly, adaptive temporal sampling is applied. Considering the different motion information between frames, if the inter-frame motion is enough smooth, some frames are skipped to obtain high compression efficiency. On the other hand, if the abrupt motion occurs, the original video sequence is up-sampled to generate a new-length video with adaptively redundant frames. Here, the latest video coding standard H.264 is employed for odd and even sub-sequences. At the decoder, if two channels can work the redundant frames will be deleted to obtain the central reconstruction. If only one channel works, side reconstruction can be accepted according to the inserted frames.

For the test video "coastguard.qcif", the proposed scheme can still consistently improve around 0.8-1.5dB in central distortion and 0.5-1.7dB in side distortion over the conventional scheme (frame splitting directly) at the bit rate from 50kbps to 300kbps per channel.

This work was supported in part by Beijing Key Laboratory of Advanced Information Science and Network Technology, NSFC (No.60903066, No. 61103113), BNSF(No.4102049), 973 program (2012CB316400), SRF-DPHE(No. 20090009120006) and JP-NSF (BK2011455).

1068-0314/12 $26.00 © 2012 IEEE
DOI 10.1109/DCC.2012.69