J. Vis. Commun. Image R. 32 (2015) 237–245

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Error concealment for the transmission of H.264/AVC-compressed 3D video in color plus depth format ☆

Wen-Nung Lie [a,b,*], Guan-Hua Lin [a]

[a] Department of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan, ROC
[b] Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI), National Chung Cheng University, Chia-Yi, Taiwan, ROC

A R T I C L E I N F O

Article history:
Received 27 April 2015
Accepted 22 August 2015
Available online 31 August 2015

Keywords:
Error concealment
Frame loss
3DTV
3D video transmission
Color plus depth
H.264/AVC
Motion vector extrapolation
Boundary matching algorithm

A B S T R A C T

An error concealment method for the transmission of 3D video in color/depth dual stream format (i.e., separately encoded) is proposed. We consider frame losses of color or depth along error-prone channels. Based on motion vector extrapolation from previous decoded frames, lost frames can be reconstructed. Specifically, the received information of one kind (color or depth) is intensively used when recovering a lost frame of the other kind. Our contributions come from classification of blocks into different kinds for different processing, joint color/depth matching, and shape-adaptive motion compensation. According to experiments, our proposed method, against traditional MV extrapolation (for color) and MV sharing (for depth) schemes, has shown distinctive gains in both color and depth. Our algorithm is effective in retaining integrity and smoothness of object boundaries, which is important to view synthesis for 3D display.

1. Introduction

3D video has been an emerging multimedia since 2010 when LCD manufacturers start to push their 3DTV products into the market. In the meantime, digital 3DTV programs were also subsequently broadcasted via satellites in Japan, Korea and China. The ATTEST (Advanced Three-dimensional Television System Technologies) and 3DTV projects have been driven in Europe for several years. These activities all lead us to a promising 3D stereo era.

H.264/AVC has prevailed for several years in compressing traditional 2D videos. MVC (Multi-view Video Coding), being an extension of H.264/AVC, enables efficient encoding of stereoscopic [1] and multi-view video captured by different cameras at different positions for the same scene. Joint Video Team (JVT) has released MVC-related reference software, known as JMVC (Joint Multi-View Video Coding). Recently, both the stereo and multi-view high profiles have been standardized, whereas another format, multi-view video plus depth (MVD), which includes information about per-pixel scene geometry for each view, has also been discussed intensively for 3D video. 3D videos of video plus depth (V+D) format are suitable for novel view synthesis, via DIBR (Depth-Image Based Rendering) technique, for auto-stereoscopic display while they are encoded and transmitted to the receiver.

Error concealment (EC) techniques [2,3] have been proposed to reduce visual quality degradation due to error propagation while communicating with videos in noisy channels. They can be categorized into: temporal or spatial. When temporal information is considered, motion vectors (MVs) of the corrupted macroblocks (MBs) would be recovered at decoder so that the result of motion compensation thereafter can be used for error replenishment. MV recovery is promising if high correlations exist in the temporal domain. On the other hand, EC techniques based on spatial interpolation/extrapolation are only considered for erroneous frames that do not have suitable reference sources in the temporal direction (e.g., the 1st frame in a sequence or the boundary frame in a scene change).

Recently, EC researches regarding 3D video have gained increasing attentions. Depending on the format of 3D information transmitted, EC techniques can be targeted at: (1) multi-view [23] or (2) MVD sequences [4–9]. The former is featured of exploiting information from other views to conceal errors found in current view, while the latter, on the other hand, explores the possibility of using one kind (color or depth) of information to aid in the error concealment of the other kind. In this paper we focus on the transmission of single-view video-plus-depth (VPD), which is a special case of MVD.

☆ This paper has been recommended for acceptance by M.T. Sun.
* Corresponding author at: Department of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan, ROC.
E-mail address: ieewnl@ccu.edu.tw (W.-N. Lie).

http://dx.doi.org/10.1016/j.jvcir.2015.08.012
1047-3203/

Regarding the compression of VPD sequences, the color and depth sequences can be separately encoded based on two independent H.264/AVC coders [6,7,9]), or jointly encoded [4,5] into a single stream. Hewage et al. [4,5] proposes to use SVC (Scalable Video Coding) layered encoding techniques for VPD sequences, where the color information is encoded as the base layer and the depth information is encoded as the enhancement layer. The enhancement layer shares or predicts [4,5,10] MVs from the base layer. Both separate and joint encoding methods have their own advantages. Joint encoding has the advantage of slice/frame synchronization and MV sharing [4] between color and depth information, but its MV sharing or prediction strategy between the base and enhancement layers might cause a degradation of coding efficiency due to less correlation coefficient (0.1–0.4 on average) [10] between them, especially for high-activity videos. On the other hand, separate coding prevents error propagation there between. Here, separate encoding of color and depth based on H.264/AVC standard is adopted in this paper.

Compressed video data are often transmitted in units of packets: (1) slice as a packet [6,9,11,12,23] and (2) frame as a packet. Slice loss is often considered in 2D video transmission [2,3]. For 3D video transmission, researchers [4,7,8] often set a slice to contain a full frame of compressed color or depth information to reduce overheads in packetization. Similarly, a packet loss in this paper will mean a whole-frame loss, either in color or depth domain. For frame loss, only temporally adjacent information is available and techniques of extrapolation from other frames can apply.

Several EC algorithms [6–9,11–13] were proposed for 3D videos of V+D format, where color and depth information are separately encoded. Liu et al. [6] and Yang et al. [11], considering slice losses, proposed a depth-assisted error concealment algorithm for the color part, e.g., incorporating depth terms in matching. In [11], depth_MVs (i.e., MV in depth video) are extracted and used to assist color error concealment based on BMA (Boundary Matching Algorithm). Doan et al. [9], considering the transmission of multi-view video plus depth sequences, proposed an enhanced temporal error concealment method by estimating the missing color_MVs (i.e., MVs in color video) of the corrupted MBs based on synthesized information obtained from correctly received or concealed depth maps. In [7], Yan et al. proposed a frame loss EC algorithm, where MV for each pixel of the lost color frame is extrapolated from the previous frame and validated with the aid of depth information. Liu et al. [13] reused color_MV for depth block encoding of the occluded regions, thus forming redundant MV information to recover the lost MVs for the occluded color blocks. Most of the afore-mentioned works [6–9,11,13] use depth information to assist in recovering MVs of the lost color MBs or pixels, whereas adopt a simple MV-sharing (or, reuse) strategy when concealing depth losses or errors. In [12], intra-coded frames of the color sequence are offset from those of the depth sequence so that temporal error concealment of the slices in color intra frames can be enhanced with the assistance of MVs extracted from the depth sequence.

In this paper, we focus on error concealment of 3D video which is of single-view V+D format, separately encoded for color and depth information, and encounters frame loss. Our contributions come from several aspects: (1) classification of blocks into different kinds ("Occ", "Disocc", and "static" in Section 3.2) for different processing (in MV refinement and extrapolation), (2) joint matching based on color and depth information, and (3) shape-adaptive motion compensation (Section 3.3.3). Based on them, recovery of lost information of one kind (color or depth) is achieved with the aid of the received information of the other kind, thus getting better recovered qualities than those of prior works. In addition to the MV extrapolation with the aid of depth information for the color part, another contribution comes from MV-sharing with a modified BMA-like process (Section 4) for the depth part. This resolves the problems or degradations when either single source (e.g., depth) is not accurate enough or the correlation between both domains is not high enough.

2. Traditional motion vector extrapolation algorithms

Here, the traditional PMVE (Pixel-based Motion Vector Extrapolation) [14] and HMVE (Hybrid Motion Vector Extrapolation) [15] are described first. PMVE, being extended from BMVE (Block-based Motion Vector Extrapolation) [16], calculates pixel-wise MVs for the lost frame. Fig. 1 illustrates the lost frame f_t, where four rectangles represent the extrapolated MB1–MB4 from f_{t-1} (based on extrapolated MVs, i.e., linearly reversing the MVs from f_{t-1} to f_{t-2}) by assuming that the motion in the video is smooth or continuous. It is seen that pixels marked with "•", "○", "■", "□" are covered with extrapolated data from (1) MB1 and MB2, (2) MB4, (3) MB2, MB3, and MB4, and (4) none, respectively. For pixels covered by at least one extrapolated MB, its MV is estimated by averaging the MVs corresponding to all overlapped MBs. For pixels not covered by any of the extrapolated MBs, its MV is duplicated from the MV of the same pixel in the previous frame.

The performance of PMVE degrades when a large object motion exists. HMVE [15] is then proposed to improve PMVE. HMVE first classifies each pixel into one of three categories, according to the number of extrapolated MBs that cover it and the 4×4 block where it resides. For each category of pixels, a set of MV candidates, including the MVs determined from PMVE, the MV possessing the largest coverage with the belonged 4×4 block, etc., is determined. Different categories have different sources of MV candidate set and different MV extraction from the set. More clearly, the final MV for each pixel in the lost frame is estimated as a compound of pixel-based and block-based MV extrapolations.

3. Error concealment for color frame loss

Since the frame loss is considered, there are two ways for MV recovery of the lost frame: (1) MV extrapolation from the previously received color frame, and (2) MV re-use from the decoded depth frame at the same time. However, it was shown that depth_MV is probably inaccurate due to less textures in depth image and might have a low correlation coefficient of 0.1–0.4 [10] with respect to the color_MV. In view of this fact, we pay more efforts in color_MV extrapolation, with the aid of depth information.

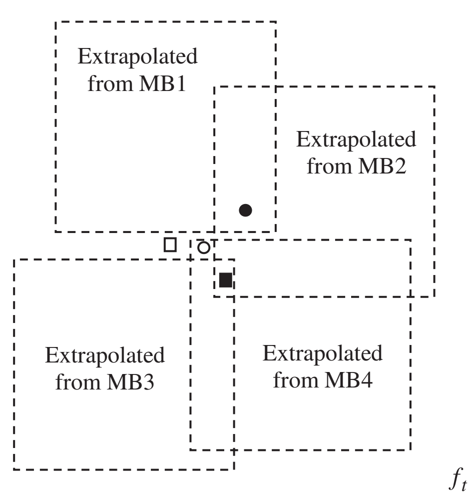

Fig. 1. The concept of PMVE.