ERROR CONCEALMENT FOR MVC AND 3D VIDEO CODING

Olgierd Stankiewicz, Krzysztof Wegner, Marek Domański

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology

ABSTRACT

In this paper we propose a novel approach to error concealment that can be applied to MVC and other 3D video coding technologies. The image content, that is lost due to errors, is recovered with use of multiple error-concealment techniques. In our work we have used three techniques: well-known temporal- and intra-based techniques and a novel inter-view technique. Proposed inter-view recovery employs Depth Image Based Rendering (DIBR), which requires neighboring views and corresponding depth maps. Those depth maps can be delivered in the bit-stream or estimated in the receiver. In order to obtain the final reconstruction, the best technique is selected locally. For that, an original recovery quality measurement method, based on cross-checking, has been proposed. The idea has been implemented and assessed experimentally, with use of 3D video test sequences. The objective and subjective results show that the proposed approach provide good quality of reconstructed video.

Index Terms— Error concealment, MVC, 3D video, depth maps, cross-checking.

1. INTRODUCTION

Modern video coding technology is very vulnerable to transmission errors due to high compression ratios and inter-frame dependences. Despite of application of sophisticated found error protection techniques, some transmission errors still corrupt video reconstructed in a decoder. A single erroneous bit results in loss of a significant portion of a slice that is often set as a whole frame. Therefore error concealment techniques are used in order to provide possibly acceptable content to the lost frames. Possibly acceptable content means an virtual frame that resembles the original as much as possible and can be used as a reference for prediction without producing visible artifacts.

There are two main approaches to error concealment for monoscopic video: intra-based and temporal-based. Intra-based techniques try to recover the lost areas with use of neighborhood from the same picture. Temporal-based approaches employ motion tracking in order to reconstruct lost fragments with use of frames that are neighboring in time.

Hitherto, very few works have been done on error concealment for MVC. In the case of multi-view coding there are more reconstruction possibilities, because neighboring views can also be used (Fig. 1).

For an example, a straightforward approach is to use inter-view correlations. In paper [1] authors employ inter-view prediction of motion vectors that are then used to recover the lost content. These motion vectors are predicted with use of disparity vectors from the neighboring available macroblocks. Unfortunately, only loss of macroblocks within a slice (not loss of the whole frame) is considered.

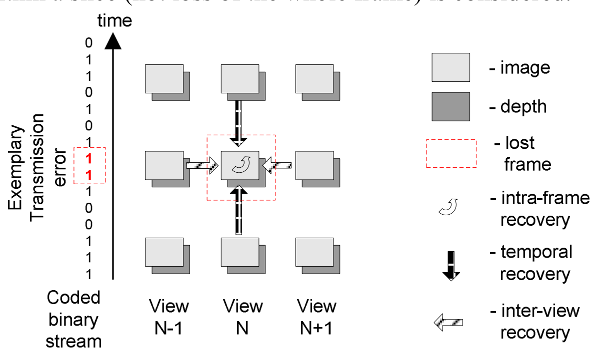

Fig. 1. Reconstruction of frame, lost due to exemplary transmission error, with use of three modes: intra-frame, temporal and inter-view, for multiple view transmission system.

Also in paper [2] motion vectors are derived for error concealment. Authors assume that both image and depth data is transmitted and employ a special scheme of motion vectors sharing between the views in order to introduce redundancy. As authors prove, such redundancy can be used for recovery of the lost frames. Unfortunately, some modifications of the coding format are required, which was shown on example of H.264/SVC coder.

In both of these abovementioned papers authors cope stereoscopic video only. A more complex case is considered in [3], which focuses on MVC. Here, advanced prediction of motion fields is employed between frames in order to reconstruct the lost frame with use of macroblock motion compensation.

To our knowledge, no work has been reported on multi-view error concealment technique that would incorporate inter-view recovery by view synthesis along with other techniques to build advanced concealment scheme. This lack was a motivation for our work. In this paper we combine such a novel inter-view recovery with intra-frame and temporal recovery modes in a new hybrid approach to multi-view error concealment problem.

2. MAIN IDEA

Error concealment technique proposed in this paper employs an original approach in which lost content is reconstructed with use of three concurrent techniques (Fig. 2): intra-frame, temporal and inter-view.

For the first two, we use some well known techniques, and as a third one we introduce a novel technique that employs Depth Image Based Rendering (DIBR). DIBR is a technique that allows synthesis of a virtual view basing on an image and corresponding depth data [4]. It is currently considered to be a core tool for future 3D video technology in works of ISO/IEC MPEG group [5].

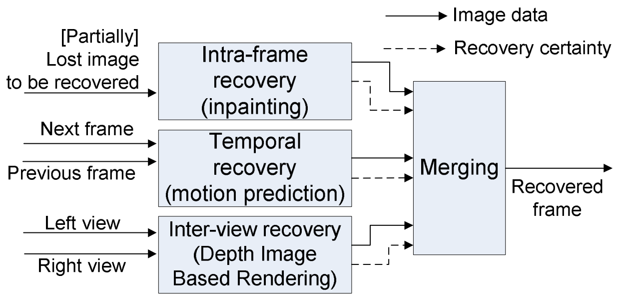

Fig. 2. Scheme of the proposed error concealment technique that employs three modes of reconstruction: intra-frame, temporal and a novel, DIBR-based, inter-view mode.

We propose that results of these techniques can be effectively merged together. It is attained by selecting the technique that locally gives the best quality of reconstruction. Because the original content is unavailable, we need another method that could provide an measurement of the quality of considered reconstructions. We propose that quality of each technique can be estimated by cross-check of its results: generated from two different reference frames (Fig. 3). The more consistent these results are, the quality of the synthesis is assumed to be higher. For temporal technique, these two references are previous and next frames, and for inter-view technique, these are frames from neighboring (left and right) views.

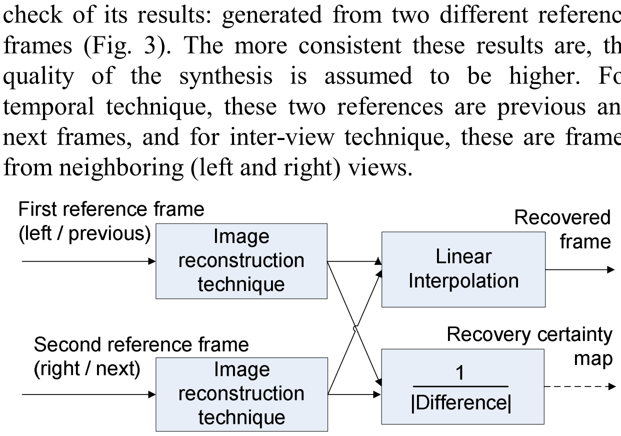

Fig. 3. General scheme of reconstruction and generation of certainty map with use cross-check.

In both cases, the certainty map $\lambda(x,y)$ is calculated basing on the following formula (1):

$$\lambda(x,y) = \frac{1}{\left| I_{first}(x,y) - I_{second}(x,y) \right|} \quad , 0 < \lambda(x,y) < 1, \tag{1}$$

$I_{first}(x,y)$ and $I_{second}(x,y)$ denote pixel luminance values at coordinates (x,y) in two images, reconstructed from the first and the second reference frames correspondingly.

In case of intra-frame recovery, where pair of reference frames cannot be used, $\lambda(x,y)$ is set to (2):

$$\lambda(x,y) = \begin{cases} 0 & if\ pixel\ is\ lost, \\ 1 & if\ pixel\ is\ available. \end{cases} \tag{2}$$

2.1. Inter-view recovery

In multiple view transmission systems, neighboring views can be exploited in order to recover the view that has been lost. We propose that such inter-view error concealment can be effectively done using Depth Image-Based Rendering. Image and depth data from left and right views (neighboring to lost one) can be used to synthesize virtual view at the position of view that is lost. Two variants of such synthesis are generated (from the left and from the right view) and then used to produce a final recovered view and recovery certainty map. This is done with use of general scheme presented in Fig. 3.

Some of the pixels might not get synthesized, e.g. those that are occluded on the left or on the right view. These pixels are left unrepaired for further processing. For those pixels, value in certainty map is set to zero.

As s synthesis tool we have used ISO/IEC MPEG group's View Synthesis Reference Software (VSRS) [4].

2.2. Temporal recovery

For temporal recovery we have used a classical approach which employs motion vectors. First, frames that are neighboring (in time) with the lost frame are found. Types of the these, depend on the type of the lost frame (see Table 1).

Table 1. Types of sought neighboring frames (previous and next) depending on GOP type of frame lost.

Type of lost frame	Type of the previous available frame	Type of the next available frame
I	I (previous GOP)	I (next GOP)
P	I (same GOP)	I (next GOP)
B	I, P or B (same GOP)	I, P or B (same GOP)

The neighboring frames, after being found, are used for macroblock motion estimation, which yields in two motion fields. These are composed of motion vectors: directing from previous frame to next frame, and from next frame to previous one (Fig. 4). Each of these motion fields is then used to predict contents of the lost frame. This is performed with well known motion compensation technique. Of course, motion vectors are accordingly weighted. Two concurrent results are generated: from the previous and from the next reference frames, which are finally used to produce a final recovered frame and a recovery certainty map (with use of scheme presented in Fig. 3).

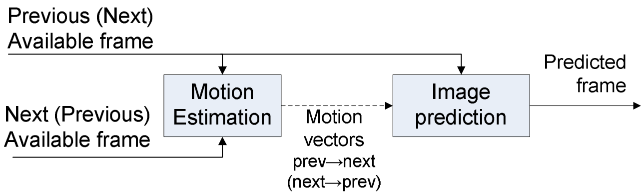

Fig. 4. Scheme of the temporal image prediction: from the previous and from the (next) neighboring reference frames.

Some of the pixels cannot be predicted from neither of motion fields, e.g. those pixels that get occluded during the motion. These pixels are not recovered and are feed for further processing with certainty set to zero.

2.3. Intra-frame recovery

Intra-frame recovery is employed on pixels that are lost due to loss of slice and in the case of two previously mentioned recovery techniques. Both inter-view recovery and temporal recovery can produce holes where reconstruction cannot be attained. This technique is also used to fill these holes.

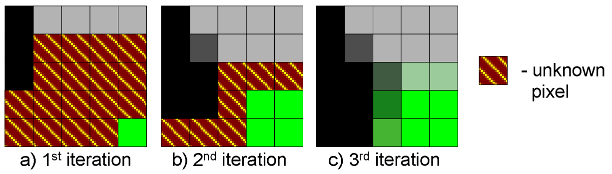

Fig. 5. Inpainting of an exemplary image with the intra-frame recovery technique.

For intra-frame recovery we employ a simple iterative flooding technique. At each iteration, each unknown pixel is recovered with average color of neighboring available pixels, from previous iteration. This process is repeated until no unavailable pixels remain (Fig. 5).

3. RESULTS

We have evaluated our error concealment method with use of standard multi-view video test sequences of ISO/IEC MPEG 3DV group. Among others, we have chosen 6 test sequences because of their quality, content diversity and depth map availability: Balloons, Kendo [6], Book arrival [7], Newspaper [8], Poznan Carpark and Poznan Street [9]. These sequences have been published for scientific use.
In order to test our error-concealment technique we have employed typical 12-frame long GOP structure: I2P3B. We assumed packet loss to be about 50% and have chosen two loss scenarios accordingly (lost frames are bold):

a) *all B frames* in GOP are lost: I**BBB**P**BBB**P**BBB**,
b) *the last P frame* in GOP is lost: IBBBPBBB**P**BBB.

Also, we assume that the lost frames are not used as a reference for prediction and thus, other frames (e.g. B-s) also have to be reconstructed (underlined). Such assumptions are severe, because typical errors lead only to desynchronization of a binary decoder until the end of a slice and thus frame with concealed errors can be successfully used as a reference. Nonetheless, in such realistic scenario artifacts would be hardly noticeable, while our aim was to reliably assess our proposal.

Sequences with such abovementioned loss of frames were put as an input for error concealment in the following configurations:

- *Motion 0* – only temporal (with no motion assumed) recovery and inpainting are performed,
- *Motion mv* – only temporal (with estimated motion vectors) recovery and inpainting are performed,
- *Full mode* – all modes of recovery (temporal, inter-view, intra-view), with certainty-based merging, are used.

Results attained with these three configurations have been assessed versus the original sequences both objectively (PSNR of luminance) and subjectively.

Subjective evaluation was following the guidelines of ITU-R BT.500 recommendation. We have used DSCQS (Double-stimulus continuous quality-scale) method, considering 95% confidence interval. The test was performed on 18 human subjects in a dark room equipped with professional 24'' LCD monitor purposed for subjective evaluation. Audience was at distance of about 2.3 meter from the screen. The whole session was composed of 31 randomized tests.

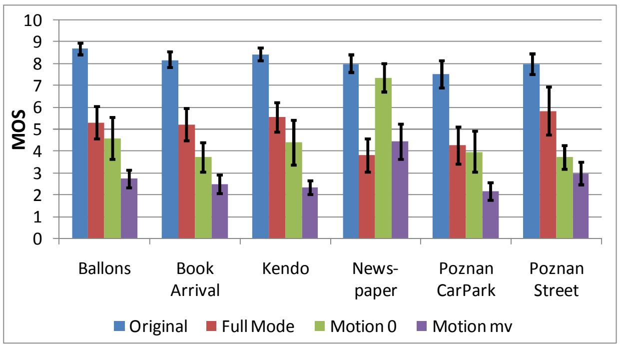

Fig. 6. Subjective evaluation results for scenario a) (all B frames lost) in continuous scale 0÷10, with 95% confidence intervals marked.

The results for scenario a) presented in Fig. 6 and 7 show that in general proposed algorithm ("*Full mode*") has the best performance of the all tested modes, although this superiority is not always statistically significant in case of subjective evaluation – Fig. 6. Also, in case of Newspaper sequence, "*Motion 0*" turns to be better than others. We believe that this results from lighting differences and some depth mismatches between left and right views in this sequence. These yield in background flickering which probably impacts final outcome of the proposed algorithm ("*Full Mode*" employs synthesis from neighboring views). Perhaps, enhancement of the depth data would lead to definite improvement of the obtained results.

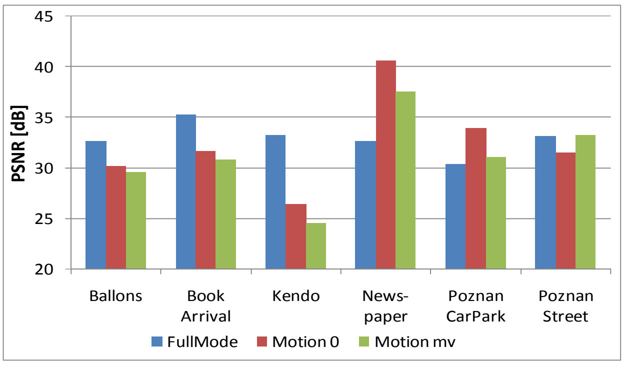

Fig. 7. Objective evaluation results for scenario a) (all B frames lost) – PSNR of luminance.

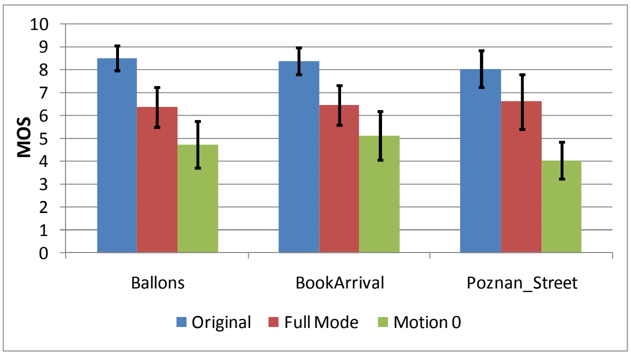

Fig. 8. Subjective evaluation results for scenario b) (the last P frame lost) in continuous scale 0÷10, with 95% confidence intervals marked.

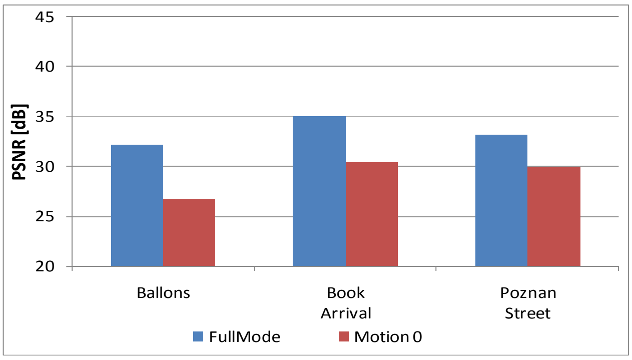

Fig. 9. Objective evaluation results for scenario b) (the last P frame lost) – PSNR of luminance.

Due to these facts, for scenario b), we have limited set of tested sequences to the ones that do not suffer from abovementioned problems. Also, we have resigned form "*Motion mv*" mode due to its low performance.

Results for scenario b) - Fig. 8 and 9 – are very similar to those obtained in scenario a). Although is not statistically proven (because of the confidence intervals overlap), in general, proposed algorithm outperforms temporal concealment technique ("*Motion 0*").

Objective results (Fig. 7 and 9) shown that quality measured by PSNR of the proposed approach is about 30÷40dB. This relatively high as for an error concealment technique.

4. CONCLUSSIONS

We have presented a novel general approach to error concealment for MVC and related 3D video coding technologies. Its application requires availability of the depth maps, which can be transmitted to the receiver or can be estimated from transmitted neighboring view.

Also an implementation that employs significant novelties has been presented: 1) an application of Depth Image Based Rendering for inter-view reconstruction and 2) estimation of quality of reconstruction based on a cross-checking.

Experimental assessment, both subjective and objective, has been performed. The results show that application of proposed method leads to good quality of reconstructed video.

5. REFERENCES

[1] X. Xiang, D. Zhao, Q. Wang, X. Ji, W. Gao, "A Novel Error Concealment Method for Stereoscopic Video Coding", *IEEE International Conference on Image Processing (ICIP),* San Antonio, Vol. 5, pp. 101-104, Oct. 2007.

[2] C. Hewage, S. Worrall, S. Dogan, A.M. Kondoz, "Frame concealment algorithm for stereoscopic video using motion vector sharing", *IEEE International Conference on Multimedia and Expo (ICME)*, Hannover, pp. 485-488, April 2008.

[3] S. Liu, Y. Chen, Y.-K. Wang, M. Gabbouj, M. M. Hannuksela, H. Li, "Frame loss error concealment for multiview video coding", *IEEE International Symposium on Circuits and Systems (ISCAS)*, Seattle, pp. 3470-3473, May 2008.

[4] M. Tanimoto, T. Fujii, K. Suzuki, "Video Depth Estimation Reference Software (DERS) with Image Segmentation and Block Matching", *ISO/IEC JTC1/SC29/WG11 MPEG/M16092*, Lausanne, Feb. 2009.

[5] Video and Requirements, "Applications & Requirements on 3D video coding," *ISO/IEC JTC1/SC29/WG11 MPEG/N11061*, Xian, Oct. 2009.

[6] http://www.tanimoto.nuee.nagoya-u.ac.jp/.

[7] I. Feldmann, M. Mueller, A. Smolic, T. Wiegand et al. „HHI Test Material for 3D Video", *ISO/IEC JTC1/SC29/WG11 MPEG/M15413*, Archamps, April 2008.

[8] Y.-S. Ho, E.-K. Lee, C. Lee, "Multiview Video Test Sequence and Camera Parameters", *ISO/IEC JTC1/SC29/WG11, MPEG/M15419*, Archamps, April 2008.

[9] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznań Multiview Video Test Sequences and Camera Parameters", *ISO/IEC JTC1/SC29/WG11 MPEG/M17050*, Xian, October 2009.