Frame Loss Concealment for H.264/AVC Stereo Video Decoders

J. Carreira, N. M. M. Rodrigues, S. M. M. de Faria, P. A. A. Assuncao
Instituto de Telecomunicações / ESTG, Instituto Politécnico Leiria
Campus 2, Morro do Lena - Alto Vieiro, 2411-901 Leiria, Portugal
jfmcarreira@gmail.com; {nuno.rodrigues, sergio.faria, amado}@co.it.pt

***Abstract*—This paper deals with reconstruction of lost frames in H.264/AVC stereo video transmitted over error-prone networks. The main goal of this research work is to develop an efficient concealment method to recover lost frames in stereoscopic video. The proposed method combines the use of both motion and disparity vectors to reconstruct missing frames in the auxiliary view of stereoscopic video. A disparity-compensated motion field is firstly used to fill the largest possible missing regions, and then the remaining unrecovered regions are reconstructed using either temporally extrapolated motion vectors or disparity vectors. A decision algorithm based on the temporally co-located base view frame is devised to efficiently adapt the proposed concealment scheme to different video sequences and different loss events. The experimental results reveal that the proposed method consistently outperforms the existing methods that only exploit the motion and inter-view similarities. The performance increases for higher bitrates and video sequences with higher motion achieving PSNR gains up to 2.61 dB.**

***Index Terms*—3D video error concealment; frame loss; joint motion-disparity compensation.**

I. Introduction

Nowadays, three dimensional (3D) video content and applications are emerging in the consumer market, broadcasting services and internet, as an extension of the existing 2D video, enhancing the user experience into a more immersive and natural viewing experience. The recent success of 3D movies and video games is seen as being partially responsible for the increasing interest in 3D visual contents, electronic devices, services and applications [1]. Furthermore, there is also a rapid development of 3D video-based technologies for acquisition, processing, coding and displaying, allowing 3D video delivery services to be available to the general public through current communication networks, either at home or in a mobile environment.

Several applications are emerging associated with the 3D video, such as the free viewpoint TV, which requires an higher number of viewpoints, also demanding for higher bitrate allocation in the transmission system. This requires efficient compression of the video information, in order to be delivered under common video channels, without significant modifications. Thus, the recent developments in the field of 3D video coding and transmission are mostly based on either the H.264/MVC standard [2] or the video plus depth format, enabled by the MPEG extension known as MPEG-C Part 3.

In generic multiview video communications, the state-of-the-art codec H.264/MVC is currently used in most stereoscopic video applications. Since this is a predictive coding format, MVC compressed streams are very sensitive to transmission errors and packet losses [3]. Whenever a frame is lost, error propagation through dependent frames increases the degradation of the reconstructed video quality and, consequently, the quality of experience. Digital transmission through networks is, in general, prone to errors and data loss. These errors can be handled in different ways, such as retransmission, error correction, and error concealment.

The main goal of this paper is to contribute with an efficient error concealment technique to cope with errors in stereoscopic video H.264/MVC streams. The concealment method developed in this work aims to recover from the loss of a whole frame. Therefore, the proposed method exploits the undamaged received information in temporally and spatially adjacent frames.

Existing concealment techniques for 2D video may be applied to stereoscopic video in the case of frame loss, *e.g.*, the direct motion-copy [4]. For instance, the method proposed in [5] is based on the assumption of constant motion, performing two-way motion vector extrapolation (*i.e.* pixel and block based) to recover the missing information. However, in stereoscopic video coding where temporal and inter-view compensation is combined, the temporal motion vector extrapolation methods may present a poor performance, since in this case it is not always possible to derive the inter-view vectors. Therefore, methods that utilise both temporal and inter-view information (vectors or an estimated disparity map) exhibit better performance. The method proposed in this paper also aims to overcome the illumination problems of some recently proposed error concealment strategies, such as those presented in [6], [7].

The method proposed in [7] relies on the motion similarity between the two stereoscopic views. The missing frames are reconstructed by using motion compensation after projecting both the motion information and the prediction residue from the error-free view, into the missing frame. Despite the problems of this method with illumination changes across views, it deals fairly well with the occlusion problems. The method

This work has been supported by IPLeiria and FCT, Portugal (Projects P3DTV and PTDC/EEA-TEL 114487/2009)

proposed in [8] mainly deals with the problem caused by illumination changes between two views during the inter-view error concealment. The method aims to recover the missing regions using the available information in the neighbour blocks.

In this paper we propose a joint error concealment method to reconstruct lost frames in stereoscopic video decoders using motion and disparity vectors, such as H.264/MVC. This method combines motion and disparity vectors to compute the motion/disparity field of the lost frame in the auxiliary view. This method extends beyond the existing ones, by estimating a disparity map for the missing instant through the extrapolated motion vectors (MVs), and by exploiting the error-free information of both base and auxiliary views. The motion field of the missing frame is determined based on the disparity compensated MVs present from the error-free view, disparity vectors fetched from the estimated disparity map, and temporally extrapolated MVs.

The paper is organised as follows: after this introduction, section II explains the proposed method to recover the lost motion field in the auxiliary view of the 3D video while the experimental results are presented in section III and section IV concludes this paper.

II. Proposed Method

In this section an error concealment method to reconstruct lost frames in stereoscopic video decoders based on H.264/MVC is proposed. In [9] the authors proposed an error concealment method that used a combined motion field to reconstruct the missing frame. This method combines motion and disparity information to estimate the motion/disparity field for the lost frame, which is used to reconstruct the missing frame itself. Since disparity vectors (DVs) may not always lead to better reconstruction of the lost frame, a decision algorithm is proposed to choose when the disparity information should be used.

It is assumed that the base view is not corrupted, *e.g.*, by using a high priority transmission channel. Figure 1 illustrates the type of information used to reconstruct the missing frame. The two stereoscopic views are shown, *i.e.*, the base view frames (f^0), the auxiliary view frames (f^1), the disparity map (D), the motion vectors (mv) and the disparity vectors (dv). In this example, the lost frame, *i.e.*, frame of instant t of view 1 (f_t^1), is represented with a dashed square. The proposed method firstly estimates a set of motion vectors $mv_t'^1$ for the missing frame f_t^1 (see 1 in Figure 1) by extrapolating the MVs, mv_{t-1}^1, associated with the last successfully decoded frame in view 1, using the following equation:

$$mv_t'^1(x,y) = -\frac{t-t_C}{t_C-t_R} \times mv_{t-1}^1(x,y), \qquad (1)$$

where t_R is the time instant of the reference frame pointed by $mv_{t-1}^1(x,y)$, and t_C refers to the time instant of the frame from which the MVs are extrapolated (*i.e.*, $t-1$).

Secondly, the disparity map of the last error-free stereo pair (2) is obtained by using the variational method described in [10]. Subsequently, a motion compensated version of D_{t-1}

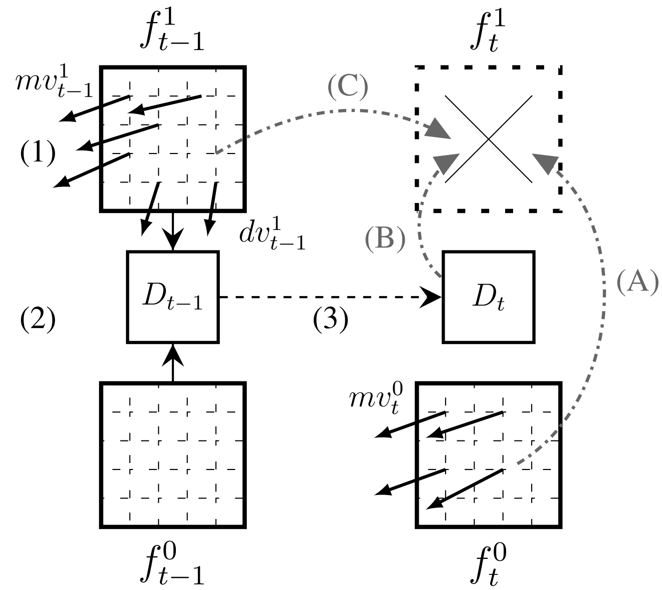


Figure 1: Frame loss and available information.

is used, as the disparity map for the missing instant. To accomplish this, the MVs previously extrapolated ($mv_t'^1$) are used. As shown in Figure 1, the last error-free stereo pair occurs at $t-1$, then the disparity map D_{t-1} is projected onto the missing instant t (3), using motion compensation together with the extrapolated MVs, determined by Equation 1. The result is the disparity map (D_t) for the instant t.

Finally, using the previously estimated disparity map, the lost motion field is reconstructed by using the MVs of the temporally co-located frame in the base view, f_t^0 (see arrow A in Figure 1). Those MVs are disparity compensated using the disparity map D_t, in order to fetch the pixels' values from the frames in the auxiliary view. Note that, each INTRA-coded block in frame f_t^0 creates a region without an associated MV in the recovered motion field. In order to fill those regions, one of the following is applied:

- The disparity map D_t is used to recover disparity vectors for the unrecovered regions of the motion field. This process is represented in the figure by arrow B. Those disparity vectors are used to fetch the blocks from the base view frame (f_t^0), decoded without errors.
- The MVs obtained from temporal extrapolation of the mv_t^1, are used to fill the gaps, represented by arrow C in the figure. Moreover, the disparity vectors (dv_{t-1}^1) present in f_{t-1}^1 are also used.

For those stereoscopic sequences characterised by small motion and high texture complexity, the difference between successive frames is small and the difference between neighbour pixels within each frame is high. In those sequences, small errors in the disparity map have a strong negative effect on the quality of the reconstructed frame, since the neighbour pixels have high differences. Therefore, when low motion is presented, the reconstruction of the lost motion field should use MVs instead of DVs.

In order to decide when disparity vectors should be used to recover the missing frame in the auxiliary view, a decision algorithm, based on the number of INTRA block and the number of non-zero MVs, is used. The decision algorithm aims to decide between the step (B) and (C) and it is applied to the whole frame. The decision is based on the factor F determined using the following equation:

$$F = 0.6 \times (INTRA - 14\%) + 0.4 \times (NZMV - 9\%)\,, \quad (2)$$

where $INTRA$ and $NZMV$ is the number of INTRA block and non-zero MVs, respectively. On the one hand, when F is larger than zero, the sequence is assumed to have higher motion and DVs provided by the disparity map are used to fill the gaps in the recovered motion field (step B). On the other hand, if F is a negative value the temporally extrapolated vectors from the frame f^1_{t-1} are used (step C). In the decision process, the values used are related with the total amount of blocks, therefore the expression is resolution independent.

III. Experimental results

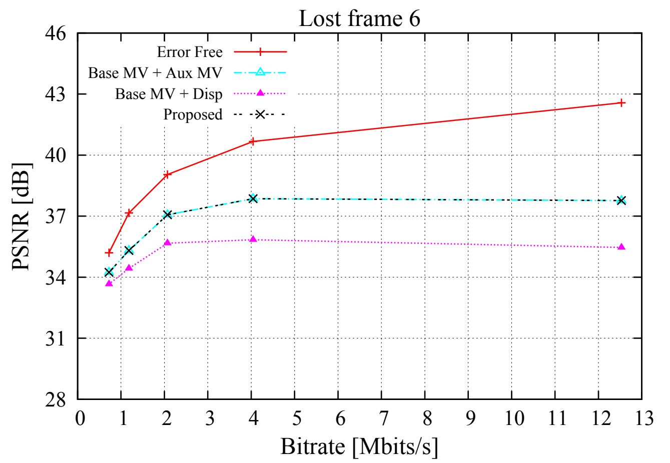
A performance evaluation of the proposed joint algorithm is presented in this section. In the experiments four well-known stereoscopic test sequences, with different resolutions and distinct types of motion and texture complexity were chosen. These sequences also represent different kinds of disparity values. Table I presents a summary of the sequences' characteristics, as well as the spatial resolution in pixels and the view IDs (*i.e.*, view number) used for the base and auxiliary view. These test sequences were captured using 1D camera array.

The test sequences were encoded using version 18.3 of the JM reference software, Stereo High profile with an IDR period of 20 frames, a GOP structure IBPBP using 2 reference frames with inter-view prediction enabled, and a range of 64 pixels was used for motion/disparity search. The available coding modes in the H.264/MVC standard were used, and the same value for the quantisation parameter (QP) was configured for I-, P- and B-slices. The sequences were enconded at 30 frames per second.

The performance of the joint motion and disparity concealment algorithm, presented in Section II, is firstly evaluated. Figure 2 presents the quality of the reconstructed frame for two different frame loss events (frame 6 and 46) in the sequence Book Arrival. Although both of the lost frames correspond to a P-frame (reference), higher motion at frame 46 is present. The results show that for higher motion, the use of DVs to fill the gaps in the reconstructed motion field (Base MV + Disp) achieves higher quality than the use of MVs available in the auxiliary view frame (Base MV + Aux MV). This indicates that in the case of high motion in the auxiliary view, disparity can provide more accurate vectors than motion. Thus, they improve the quality of the reconstructed frame when combined with the available motion vectors of the base view frame. In contrast, for low motion, the reconstruction of the lost motion field using the auxiliary view MVs results in higher quality, as shown in the Figure 2, when frame 6 is lost. However, the

Table I: Stereoscopic video sequences used to evaluate the performance of the proposed concealment technique.

Sequence	Resolution	Description	Views used
Akko & Kayo	640×480	High translational motion, with two moving persons; moderate texture details; moderate depth.	50/49
Book Arrival	1024×768	Moderate translational motion; moderate depth values; moderate texture complexity.	9/8
Kendo	1024×768	High translational motion with two moving persons; moderate depth; moderate texture.	3/2
Champagne Tower	1024×768	Low motion; high depth; complex texture with transparent objects.	40/39



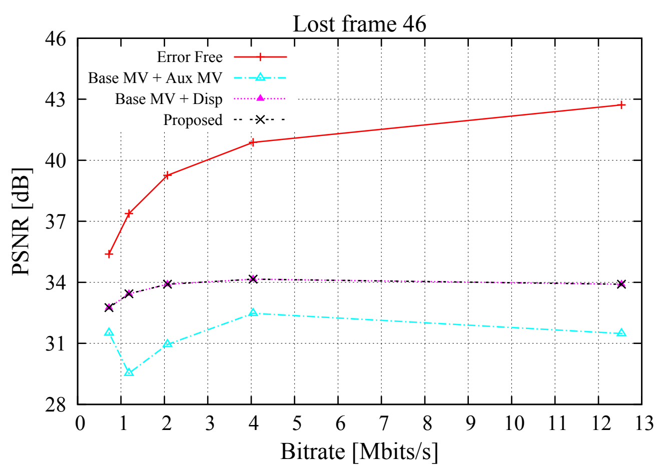


Figure 2: Quality of the reconstructed frame for two lost events in the sequence Book Arrival.

proposed method, is able to achieve the same quality as the best of the two approaches, revealing the relevance of using a decision method in the concealment process.

In a second set of experiments, the proposed method is evaluated against three other methods: frame-copy (FC), that replaces the missing frame with the previously decoded frame in the same view; the motion-copy method (MC), implemented

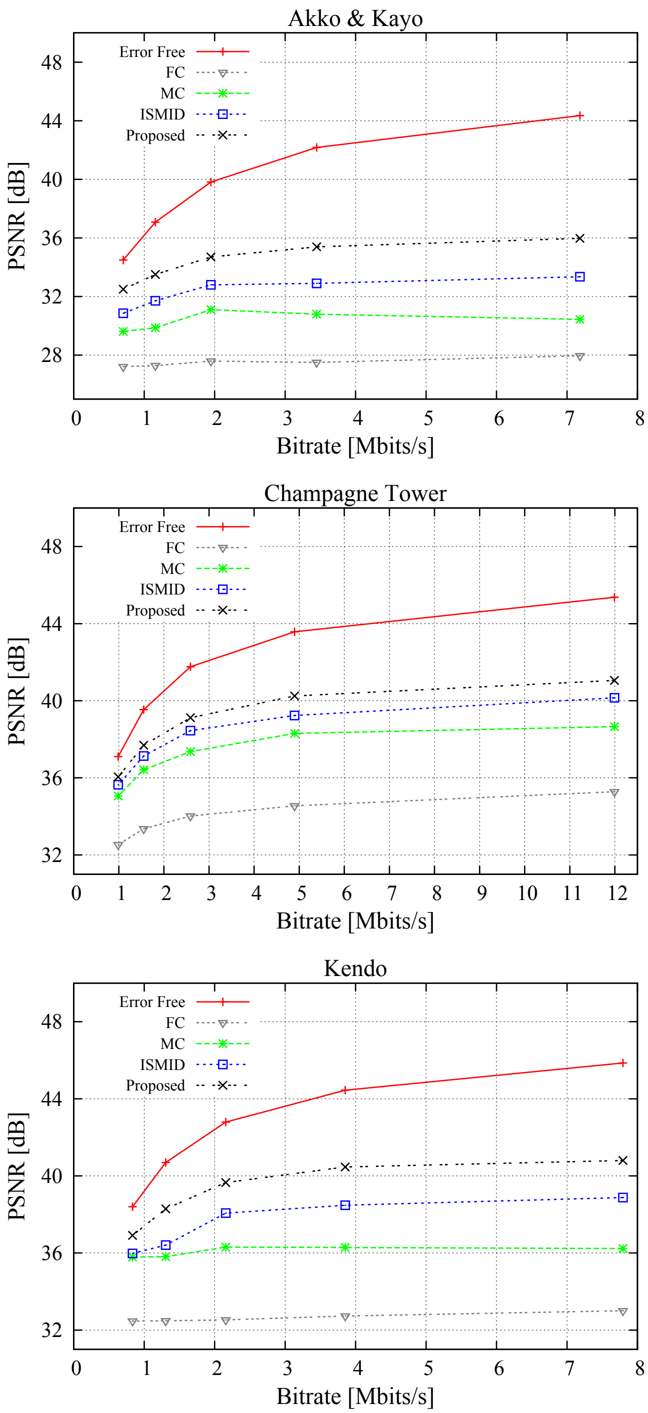


Figure 3: Average quality evaluated with the PSNR under 10% of frame loss.

in this work for the auxiliary view in the JM reference software, as proposed in [4]; the method based on the inter-view similarities (ISMID) proposed in [7].

Figure 3 shows the average quality results, evaluated with the PSNR, for different sequences, when the following frames are lost: 6, 16, 26, 36, 46, 56, 66, 76, 86 and 96. This corresponds to drop two frames between each IDR frame. The results show that the proposed method is able to outperform all other methods tested for different stereoscopic sequences. One should notice that, the sequence Champagne Tower presents lower motion than the sequence Kendo (see Table I), and the results show that higher gains for the sequence Kendo are achieved. Thus, we can conclude that the efficiency of the proposed method increases with the sequences' motion. Moreover, the analysis of the average quality results also reveals that the advantage of the proposed method increases for higher bitrates, achieving a maximum PSNR gain of 2.61 dB for sequence Akko & Kayo @ 7.18 Mbits/s. These simulation results demonstrate the usefulness of combining motion and disparity vectors to recover missing frames in 3D video.

IV. Conclusion

In this paper an efficient joint motion-disparity concealment method for 3D video decoders was described and assessed. The results revealed the advantage of the proposed method under different conditions, such that, it can be used as an efficient alternative to motion-copy or to more complex methods, that exploit inter-view dependencies. These results show that DVs can be used to provide more accuracy to the reconstructed motion/disparity field under high motion scenarios, achieving higher quality than using only temporal extrapolated MVs from the auxiliary view. Therefore, a decision algorithm was proposed, to accurately use DVs in the concealment process, improving the quality of the reconstructed frame. The overall results and conclusions presented in this paper provide relevant knowledge to deal with missing frames in the auxiliary view of 3D video under different conditions.

References

[1] P. Merkle, K. Müller, and T. Wiegand, "3D video: acquisition, coding, and display," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 946–950, May 2010.

[2] A. Vetro, T. Wiegand, and G. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[3] B. Micallef and C. Debono, "Error concealment techniques for multi-view video," in *IFIP Wireless Days (WD)*, Oct. 2010, pp. 1–5.

[4] S. K. Bandyopadhyay, Z. Wu, P. Pandit, and J. M. Boyce, "An error concealment scheme for entire frame losses for H.264/AVC," in *Sarnoff Symposium, IEEE*, Mar. 2006, pp. 1–4.

[5] B. Yan and H. Gharavi, "A hybrid frame concealment algorithm for H.264/AVC," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 98–107, Jan. 2010.

[6] Y. Chen, C. Cai, and K.-K. Ma, "Stereoscopic video error concealment for missing frame recovery using disparity-based frame difference projection," in *IEEE International Conference on Image Processing*, Nov. 2009, pp. 4289–4292.

[7] T.-Y. Chung, S. Sull, and C.-S. Kim, "Frame loss concealment for stereoscopic video based on inter-view similarity of motion and intensity difference," in *IEEE International Conference on Image Processing,*, Sep. 2010, pp. 441–444.

[8] H.-Y. Yang, J.-Y. Lin, H.-W. Tsao, and Y.-C. Fan, "Algorithm and architecture design of illumination changes adaptive motion estimation," in *IEEE International Symposium on Consumer Electronics*, May 2009, pp. 565–568.

[9] J. Carreira, P. Assuncao, N. Rodrigues, and S. Faria, "Frame loss concealment for 3D video decoders based on disparity-compensated motion field," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, Oct. 2012, pp. 1–4.

[10] S. Kosov, T. Thormählen, and H.-P. Seidel, "Accurate real-time disparity estimation with variational methods," in *Advances in Visual Computing*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2009, vol. 5875, pp. 796–807.