This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

Temporal and Inter-view Consistent Error Concealment Technique for Multiview plus Depth Video

Shadan Khattak, Thomas Maugey, Raouf Hamzaoui, Shakeel Ahmad, and Pascal Frossard

***Abstract*—Multiview plus depth (MVD) is an emerging video format with many applications, including 3D television and free viewpoint television. During broadcast of compressed MVD video, transmission errors may cause the loss of whole frames, resulting in significant degradation of video quality. Error concealment techniques have been widely used to deal with transmission errors in video communication. However, the existing solutions do not address the requirement that the reconstructed frames be consistent with neighbouring frames, i.e., corresponding pixels have consistent color information. We propose a new consistency model for error concealment of MVD video that allows to maintain a high level of consistency between frames of the same view (temporal consistency) and those of neighbouring views (inter-view consistency). We then propose an algorithm that uses our model to implement concealment in a consistent way. Simulations with the reference software for the Multiview Video Coding project of the Joint Video Team (JVT) of the ISO/IEC MPEG and ITU-T VCEG show that our method outperforms benchmark techniques, including a baseline approach based on the Boundary Matching Algorithm, with respect to both reconstruction quality and view consistency.**

***Index Terms*—Multiview plus Depth, Multiview Video Coding, Error Concealment.**

I. Introduction

In video broadcasting, video data is compressed and transmitted to the home over satellite, cable, or terrestrial delivery channels. Because modern video compression schemes use entropy coding and inter-frame coding, a single transmission error can lead to error propagation that may affect several frames. To protect the transmitted data against transmission errors, video broadcasting systems typically use forward error correction (FEC). However, FEC cannot guarantee perfect recovery of the transmitted data. In mobile applications, for example, the error rate after FEC may be significant under realistic conditions [1]. For this reason, FEC is often used in conjunction with error concealment at the decoder, which aims at masking the effect of residual transmission errors.

In this paper, we study error concealment in the context of MVD [2] video. As in [3], [4], [5], the multiview texture and depth videos are compressed independently of each other using the Multiview Video Coding (MVC) standard [6] with a typical prediction structure (Fig. 1). The texture data and depth maps are encapsulated into separate packets and broadcast over an error-prone channel. Due to the high compression efficiency of MVC, it is possible to compress and encapsulate a complete frame of a low resolution sequence into a single packet. Thus, the loss of a single packet can result in the loss of a complete frame. Even when a single frame is encapsulated into multiple packets, whole frame loss is still highly probable for two reasons: (i) burst losses causing the loss of multiple successive packets are very common in video broadcast [7], [8] and are likely to corrupt all the packets of one frame, (ii) many decoders discard the full video frame even if a single packet containing parts of the video frame data is lost [9]. Hence, we assume that transmission errors lead to the loss of complete frames, such that efficient error concealment techniques are required to reduce their effect on the decoded video quality.

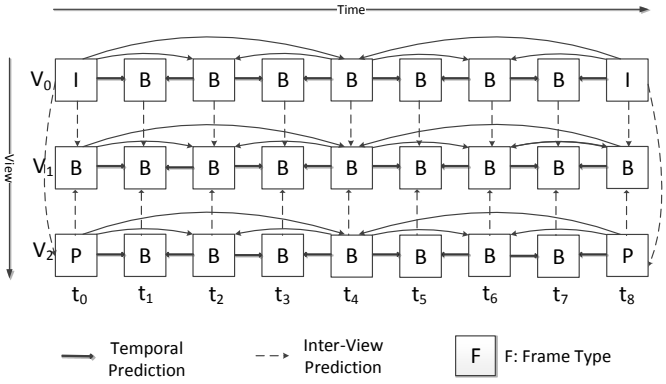


Fig. 1. Typical MVC prediction structure with three views V_0, V_1, V_2.

For multiview video, it is not only important for the concealment technique to reconstruct the individual frames with high fidelity but also to preserve the consistency between neighbouring frames, i.e., corresponding pixels in neighbouring frames (of the same view as well as neighbouring views) should have consistent color information. In most 3D applications, frames are not viewed independently. Consequently, inconsistent frames can lead to an inconsistent reconstruction of 3D scenes, which may negatively affect the viewing experience. However, the consistency requirement is ignored in existing error concealment methods [10]-[15].

S. Khattak is with the Department of Electrical Engineering, COMSATS Institute of Information Technology, Abbottabad, Pakistan (email: shadankhattak@ciit.net.pk).

T. Maugey is with Centre INRIA Rennes, 35042 Rennes Cedex, France (email: thomas.maugey@inria.fr).

R. Hamzaoui, and S. Ahmad are with the Faculty of Technology, De Montfort University, The Gateway, Leicester, LE1 9BH, United Kingdom (email: rhamzaoui@dmu.ac.uk, sahmad@dmu.ac.uk).

P. Frossard is with the Signal Processing Laboratory (LTS4), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (email: pascal.frossard@epfl.ch).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

We address this fundamental problem by proposing a scene-consistent error concealment method for MVD videos. We first introduce a novel metric for consistent 3D video reconstruction. We then exploit inter-view and intra-view correlations, as well as the geometry of MVD frames to build a set of candidate blocks for error concealment. Next, we use our consistency-based metric to select the best candidate blocks for concealment. Experimental results show that, compared to previous approaches, including [12], [13], and a baseline technique based on the Boundary Matching Algorithm (BMA) [16], our method can reconstruct the lost frames with higher fidelity while maintaining a high level of consistency between frames of the same view (temporal consistency) and those of the neighbouring views (inter-view consistency). In particular, our method can reduce flickering artefacts in 3D videos, which are often caused by inconsistencies in video frames.

The remainder of the paper is organized as follows. In Section II, we review existing error concealment techniques for 2D and 3D video. In Section III, we introduce our metric and our scene-consistent error concealment method. In Section IV, we present simulation results using the JMVC 8.5 reference software [17] for MVC. Finally, in Section V, we give our conclusions and suggest future work.

II. Related Work

In this section, we review existing error concealment methods for 3D video. We also briefly overview error concealment methods for 2D video as many error concealment methods for 3D video are extensions of 2D ones.

For 2D video based on H.264/AVC [18], a motion vector extrapolation (MVE) [19] based hybrid motion vector extrapolation method is proposed in [20]. This method considers extrapolated motion vectors (MVs) at both pixel and block levels and discards inaccurate MVs on the basis of their Euclidean distances from other MVs in the selected set of candidate MVs. Ji, Zhao, and Gao [21] and Guo et al. [22] propose error concealment methods for scalable video coding (SVC) [23]. The method proposed in [21] is based on the temporal direct mode which is usually used in regions with slow or no motion. Thus, for content with fast motion or complex texture, it might not be as efficient. Guo et al. [22] propose Intra-layer and Inter-layer concealment methods. The Intra-layer methods use the information of the same spatial or quality layer to conceal a lost frame while the Inter-layer methods use the information of the base layer to conceal a lost frame from one of the enhancement layers. While these methods may be extended to recover lost MVD frames, they do not address the issue of inconsistencies in the recovered frames.

Song et al. [10] propose three error concealment methods for MVC: temporal bilateral error concealment, inter-view bilateral error concealment, and multihypothesis bilateral error concealment. The first method uses spatio-temporal correlations in each view, the second uses inter-view correlation, while the third recovers the motion and disparity vectors of the lost block using the block matching principle [24]. For block losses in video plus depth format, Liu, Wang and Zhang [11] jointly consider the depth and neighbouring spatial and temporal information to recover the lost MVs for the corrupted blocks. The application of these methods is limited to the scenario of block losses since they depend on the availability of correctly decoded neighbouring MBs from the same frame as that of the lost MBs.

Among the methods proposed for whole-frame loss concealment in 3D video, inter-view motion vector correlation of MVC is exploited in [12]. This method first estimates the overall disparity between corresponding frames from neighbouring views. If a frame in one view is lost, its corresponding MBs are identified in a neighbouring view using the overall average disparity. The MVs of the corresponding MBs are then used to reconstruct the lost frame. This method assumes that global disparity is a good approximation of local disparities. This may not always be true and hence the efficiency of the method generally decreases as the difference between global and local disparities increases. For frame losses in stereo plus depth format, Chung, Sull and Kim [13] use a 3D image warping technique to determine matching pixels between neighbouring views and do the reconstruction based on the similarities of the motion vectors and the intensity differences of matching pixels. Hewage et al. [14] propose to share motion vectors between the texture and depth videos if a frame from either of them is lost. This method might not be very efficient when a frame contains objects with different textures placed at the same depth. Similarly, for frame losses in V+D format, Yan and Zhou [15] propose to use depth differences as a measure of the reliability of the MVs in a set of candidate MVs.

Generally, all the above methods involve the following two steps: 1. Extract several candidates for error concealment. 2. Use an evaluation criteria to discard less likely candidates and select the final candidate. The first step is non-trivial in both block and frame loss methods. The second step is even more complicated. Block based methods usually use some extension of BMA [16], which finds the difference between the outer boundary pixels of the available neighbouring blocks and the inner boundary pixels of the concealed block, while frame loss methods are usually based on simple heuristics such as the maximum overlap method [20] in case of MVE. In such methods, the MVs extrapolated from the pixels in the previous frame may not be accurate, i.e., some MVs are likely to be wrongly extrapolated, especially in large motion scenes. Another problem with these methods is that they only aim to recover the content of the lost frame without taking into consideration the effect on the consistency between the frames. Consequently, the consistency between spatio-temporally neighbouring frames that represent the 3D scene might be affected. Scene consistency in 3D video has been studied in the context of seam carving [25], image segmentation [26], feature points detection [27], and view synthesis [28], [29]. To the best of our knowledge, it has not been applied to the error concealment problem. For whole frame losses in 3D video, it is desirable to have a cost function for selecting candidate data that can efficiently conceal the lost frames by recovering the content of the lost frames with high consistency between their inter-view and temporal neighbours and hence provide a consistent viewing experience to the users.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

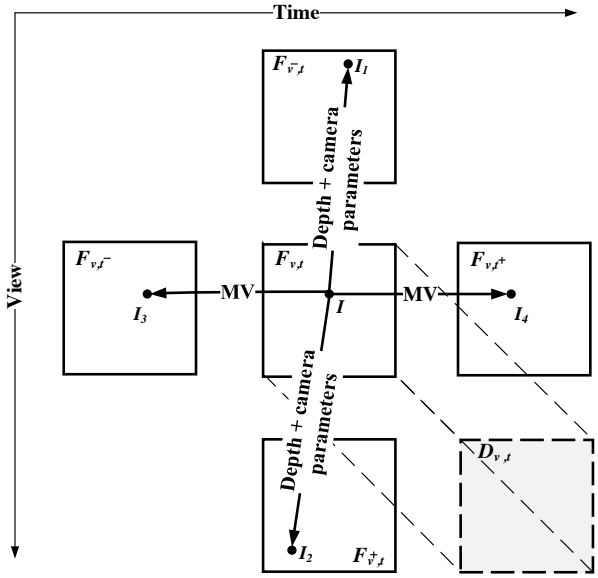


Fig. 2. Proposed scene consistency model. The pixel values I, I_1, I_2, I_3, and I_4 represent $F_{v,t}(i,j)$, $F_{v^-,t}(i_1,j_1)$, $F_{v^+,t}(i_2,j_2)$, $F_{v,t^-}(i_3,j_3)$, and $F_{v,t^+}(i_4,j_4)$ respectively.

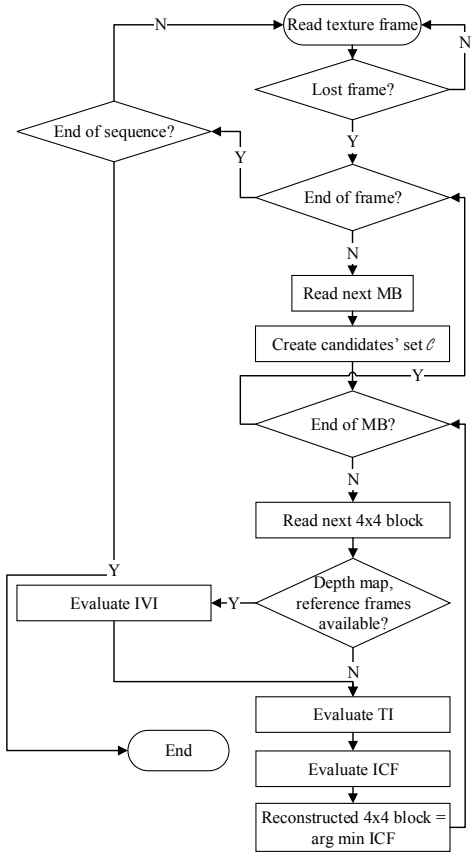


Fig. 3. Flowchart of the proposed scene-consistent error concealment algorithm which uses our consistency metric (ICF) to choose between candidate blocks to reconstruct blocks of the lost frame. In each frame, the macroblocks and the 4x4 blocks are scanned in raster order.

III. Scene consistent error concealment

In this section, we propose a new scene consistent error concealment technique, which uses the inter-view, temporal and geometric information of the neighbouring texture as well as depth frames to recover the lost frames with high consistency in MVD sequences.

A. Preliminaries

A typical MVD setup is illustrated in Fig. 2. Frame $F_{v,t}$ has two temporal neighbours F_{v,t^-} and F_{v,t^+}, two view neighbours $F_{v^-,t}$ and $F_{v^+,t}$, and a depth frame $D_{v,t}$. The intensity of pixel (i,j) in frame $F_{v,t}$ is denoted by $F_{v,t}(i,j)$.

Using 3D warping [30], we can associate to (i,j) pixels (i_1,j_1) and (i_2,j_2) in frames $F_{v^-,t}$ and $F_{v^+,t}$, respectively. 3D warping uses the depth value $D_{v,t}(i,j)$ corresponding to (i,j), the intrinsic matrices $A(v)$, $A(v^+)$ and $A(v^-)$ and the translation vectors $T(v)$, $T(v^+)$ and $T(v^-)$ of views v, v^+ and v^-, respectively and the rotation matrix $R(v)$ of view v. The intrinsic matrix $A(u)$ for view u represents the transformation from the camera coordinate system of view u to its image coordinate system while a translation vector $T(u)$ and a rotation matrix $R(u)$ describe the displacement of the camera from the origin and the direction of the camera, respectively [31]. Using these quantities, pixel (i,j) in $F_{v,t}$ is first projected into world coordinates $[u,v,w]$ via

$$[u,v,w] = R(v).A^{-1}(v).[i,j,1].D_{v,t}(i,j) + T(v). \quad (1)$$

Next, the world coordinates are mapped onto the target coordinates $[i',j',k']$ of the frame in a target view, v', via

$$[i',j',k'] = A(v').R^{-1}(v).[u,v,w] - T(v'). \quad (2)$$

Finally, to obtain pixel (i',j') (where (i',j') represents (i_1,j_1) and (i_2,j_2) when $v' = v^-$ and $v' = v^+$, respectively), the target coordinates are converted to an homogeneous form, i.e., $(i',j') = (i'/k', j'/k')$. Collectively, we call the intrinsic, translation and rotation matrices camera parameters. The warping method described in this section is normally used for view synthesis. In our paper, we exploit it to define an inter-view inconsistency metric (see the role played by (i_1,j_1) and (i_2,j_2) in Section III-B).

B. Scene consistency model

In this section, we introduce our consistency model and use it for error concealment. Our model consists of two parts: (i) inter-view consistency and (ii) temporal consistency. We define the inter-view inconsistency (IVI) at position (i,j) of $F_{v,t}$ as

$$\begin{aligned} IVI(i,j) = & \\ & |F_{v,t}(i,j) - F_{v^-,t}(i_1,j_1)| \\ & + |F_{v,t}(i,j) - F_{v^+,t}(i_2,j_2)| \end{aligned} \quad (3)$$

where positions (i_1,j_1) and (i_2,j_2) in frames $F_{v^-,t}$ and $F_{v^+,t}$, respectively, are obtained using the 3D warping method

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

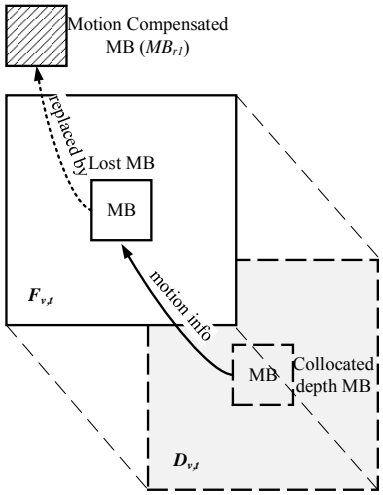


Fig. 4. Depth Motion Vector Sharing (DMS) used to create candidate MB_{r_1}.

explained in the previous section. In order to obtain high inter-view consistency, the intensity values $F_{v,t}(i,j)$, $F_{v^-,t}(i_1,j_1)$, and $F_{v^+,t}(i_2,j_2)$ should be similar.

Similarly, we define the temporal inconsistency (TI) at position (i,j) in $F_{v,t}$ as

$$TI(i,j) = |F_{v,t}(i,j) - F_{v,t^-}(i_3,j_3)| + |F_{v,t}(i,j) - F_{v,t^+}(i_4,j_4)| \quad (4)$$

where positions (i_3,j_3) and (i_4,j_4) in frames F_{v,t^-} and F_{v,t^+}, respectively, are obtained by using the motion vector MV associated with the block in $F_{v,t}$ which contains pixel (i,j) (Fig. 2), i.e., $(i_3,j_3) = (i+MVx, j+MVy)$ and $(i_4,j_4) = (i-MVx, j-MVy)$. Objects usually move with a regular motion. So if a motion vector can be used to trace an object in a past frame, the same motion vector can be used to trace the object in a future frame as well ([12]). In order to have high temporal consistency, the intensity values $F_{v,t}(i,j)$, $F_{v,t^-}(i_3,j_3)$ and $F_{v,t^+}(i_4,j_4)$ should be similar.

Finally, we combine IVI and TI into a cost function, which we call the Inconsistency Cost Function (ICF) and define as

$$ICF(i,j) = \alpha.IVI(i,j) + (1-\alpha).TI(i,j) \quad (5)$$

where $\alpha \in [0,1]$ is a weighting factor. This cost function is used to select the best blocks in the concealment method in order to maximize consistency.

To reconstruct an MB in a lost frame, the receiver first defines a set $\mathcal{C}$ of MBs from the available frames (see Section III-C). Then each 4×4 block in a MB of the lost frame is reconstructed as the 4×4 block B at the same location in an MB of $\mathcal{C}$ that minimizes $\sum_{i,j \in B} ICF(i,j)$. That is, our concealment technique picks among the candidates the most consistent one.

A flowchart of the algorithm is shown in Fig. 3.

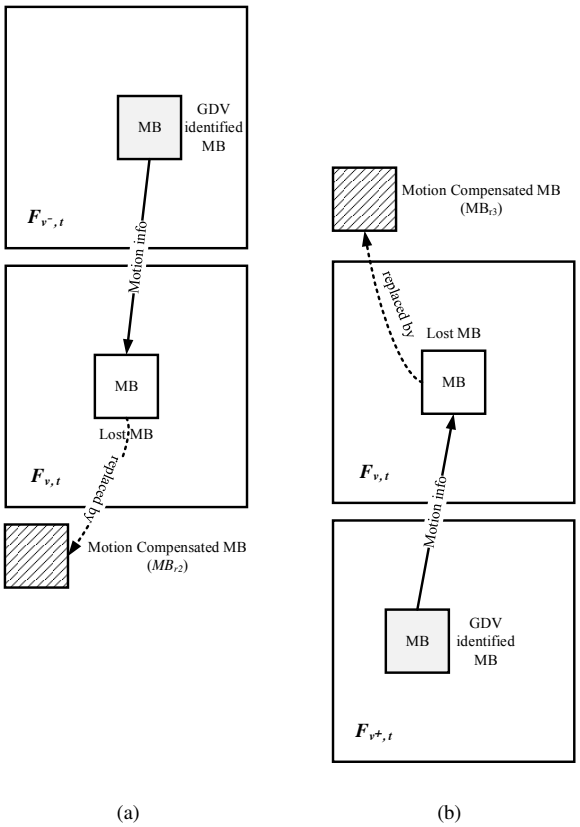


Fig. 5. Candidates (a) MB_{r_2} and (b) MB_{r_3} are formed using the MVs of the MBs identified with the Global Disparity Vector (GDV) between the current view and (a) the left view and (b) the right view.

C. Candidate MBs for reconstruction

Our concealment method considers a set $\mathcal{C}$ of four candidate macroblocks defined as follows. The first candidate, MB_{r_1}, is built by using the MVs of the collocated MB in the corresponding depth frame $D_{v,t}$ as in [14]. We call this method Depth Motion Vector Sharing (DMS) (Fig. 4).

The next two candidates MB_{r_2} and MB_{r_3} are obtained as in [12] by using the MVs of the MBs in frames $F_{v^-,t}$ and $F_{v^+,t}$ identified by using the global disparity [32] between the current view and the respective left and right views (Fig. 5).

The last candidate, MB_{r_4}, is constructed with view-synthesis [30]. We first create a synthesized version of the lost MB using the left reference frame $F_{v^-,t}$ and its corresponding depth frame $D_{v^-,t}$. We then create a second synthesized version of the lost MB using the right reference frame $F_{v^+,t}$ and its corresponding depth frame $D_{v^+,t}$. Finally, we merge the two synthesized versions such that the holes in one version are filled using the texture from the other. This fills up most of the large holes. To fill the remaining small holes, we use the morphological close operation. We call this method View-synthesis Concealment (VSC) (Fig. 6).

IV. Simulation Results

In this section, we present simulation results for our method and for three other 3D concealment techniques for whole

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

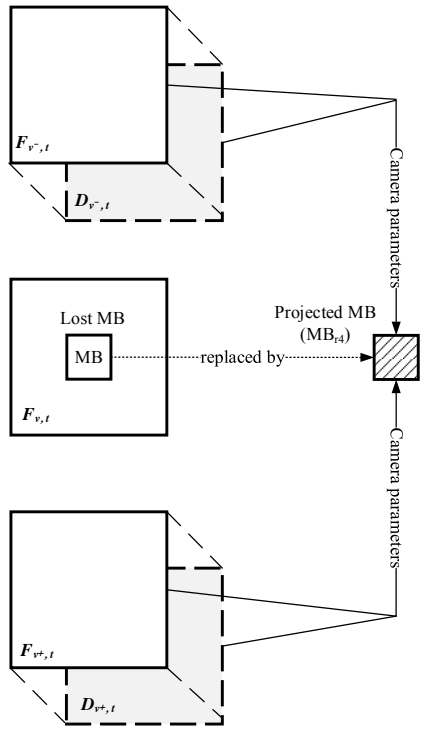


Fig. 6. View Synthesis Concealment (VSC) used to create candidate MB_{r_4}.

frame loss: the method of Liu et al. [12], the method of Chung, Sull, and Kim [13] (both described in Section II), and Boundary Matching Concealment (BMC), which we describe below. The goal of the experiments is to compare the effectiveness of the methods in the reconstruction of whole frames lost from V_1. We focus on the middle view since frames from the other two views can be recovered with conventional 2D concealment techniques (e.g., [19] and [20]).

In our method, the weighting factor α in (5) was set to 0.5. Moreover, the inter-view consistency function $IVI(i,j)$ was only partially computed if either $F_{v^-,t}$ or $F_{v^+,t}$ was lost and not used if both $F_{v^-,t}$ and $F_{v^+,t}$ or $D_{v,t}$ were lost. An alternative would have been to exploit conventional 2D concealment algorithms to recover any lost frame needed for the computation of $IVI(i,j)$.

Like our method, BMC uses 4×4 blocks from MB_{r_1}, MB_{r_2}, MB_{r_3}, and MB_{r_4} as candidate blocks but selects one of them with a slightly modified version of BMA [16]. BMA uses the difference between the boundary pixels of the lost and a concealment block to evaluate the quality of concealment. It is commonly used for recovering a lost block for which spatially neighbouring left, right, top and bottom blocks are available. In the frame loss scenario considered in our paper, these blocks are not available, so we create the first row and the first column of blocks of the lost frame using DMS. Each of the remaining blocks is recovered by finding the difference between the outer boundary pixels of its left and top blocks and the inner boundary pixels of each of the candidate reconstructed blocks. The candidate reconstructed block for which such a difference is the smallest is chosen for concealment of the current block.

We used the JMVC 8.5 reference software [17] to encode three texture views and their associated depth maps of four standard video sequences (1024 x 768 Ballet [33], 1024 x 768 Breakdancers [33], 1920 x 1088 Poznan_Street [34], and 1920 x 1088 Poznan_Hall2 [34]). 100 frames of each texture and depth view of the test sequences were used. For all sequences, each frame consisted of one slice, the frame rate was 25 frames per second, and the GOP size was 12. The quantization parameter (QP) was set to 28 for the texture and to 20 for the depth maps.

In the first experiment, we use a simple frame loss scenario where one frame from V_1 is lost to validate the consistency model.

To evaluate temporal consistency, we study the difference between the reconstructed frame and its temporal left and right neighbours. Fig. 7 shows that our method achieves better temporal consistency compared to the method of Chung, Sull, and Kim [13]. In particular, the latter method is not efficient when it faces a complex texture or object boundaries (e.g., the outlines of the subjects in Fig. 7).

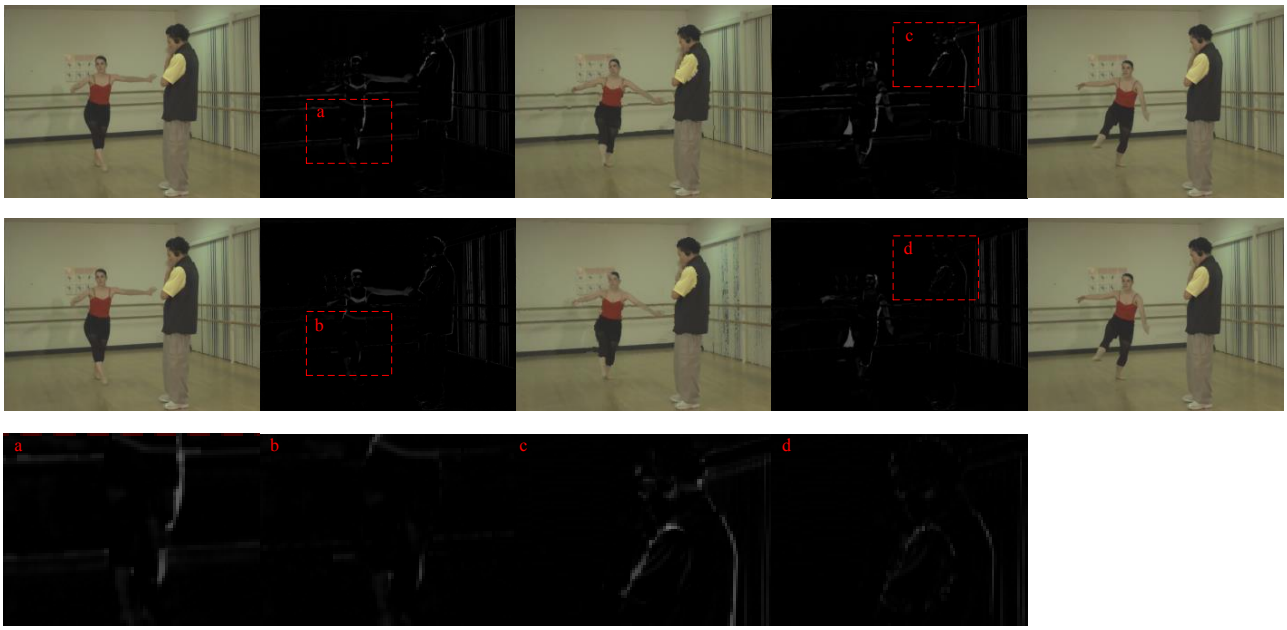
To evaluate inter-view consistency, we study the difference between the projection of the reconstructed frame in its view-neighbouring left and right frames and the view-neighbouring left and right frames, respectively. Fig. 8 and Fig. 9 show that our method achieves better view consistency compared to the method of Chung, Sull, and Kim [13].

Fig. 10 compares the reconstruction quality of our method to that of the approaches in [12] and [13]. The zoomed parts of the frames highlight the gains of our method. The gains are particularly visible for the Poznan_Hall2 sequence, which has a smaller baseline distance between cameras.

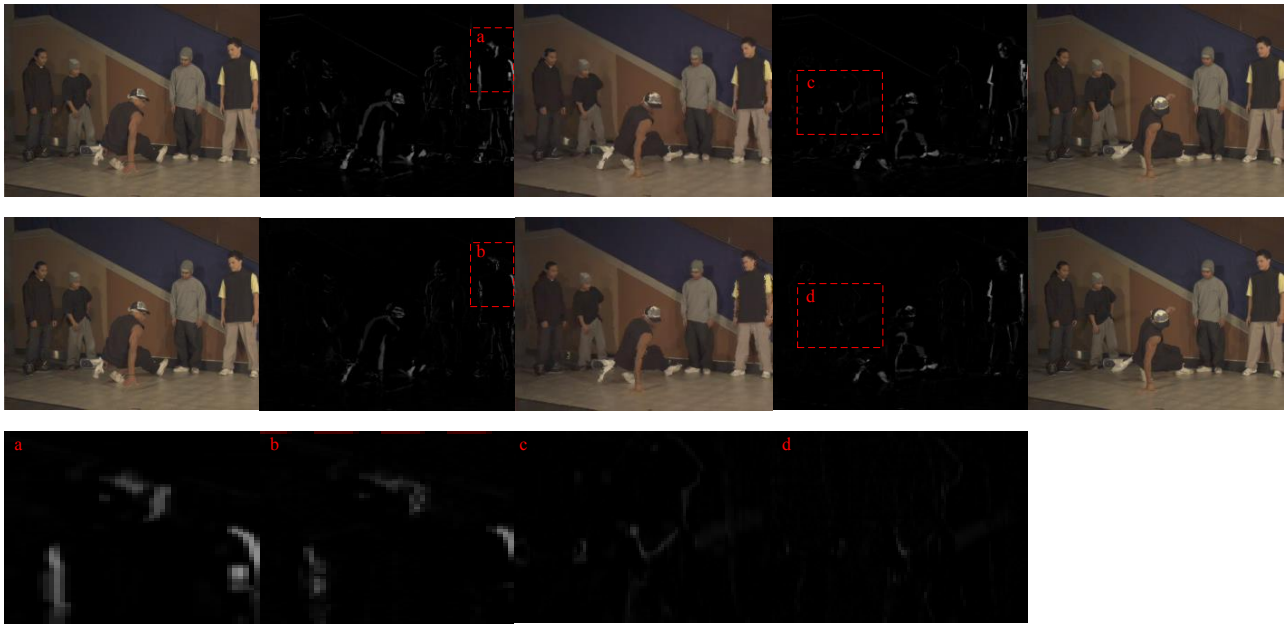
In the second experiment, we compare the PSNR performance of our method to that of the three benchmark techniques when frames (both texture and depth) from all three views are lost according to an i.i.d. process. We consider: (i) the average PSNR of the reconstructed frames (Table I), (ii) the average PSNR of all frames (Table II), and (iii) the average PSNR of the reconstructed frames and the frames that depend on them (Table III). In addition to average PSNR results, we also show the PSNR as a function of the frame number (Fig. 11). The transmission order was V_0, V_2, V_1. Texture frames were transmitted before depth frames. The simulations were repeated 50 times. If a frame in V_0 or V_2 is lost, it is not exploited for the concealment of frames in V_1. Similarly, if the required motion information is not available, a zero motion vector is used in the concealment algorithm.

The results show that our method can reconstruct lost frames with higher fidelity than the other approaches. In particular, (i) our method has the highest average PSNR, (ii) as the Frame Loss Rate (FLR) increases, so does the gain of our approach, (iii) as the distance between cameras decreases (from 20 cm for Breakdancer and Ballet sequences to 13.5 cm for Poznan_Street and Poznan_Hall2 sequences), the gain of our approach increases. The improved PSNR performance of our approach compared to BMC can be mainly attributed to the fact that BMC always relies on the decoded left and top blocks.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology



(a) Ballet



(b) Breakdancer

Fig. 7. Comparison of temporal consistency. Top: [13], Middle: our method, Bottom: Zoomed difference images. The top two rows show (from left to right): frame $F_{v,t-}$, the difference of the reconstructed frame $F_{v,t}$ and frame $F_{v,t-}$, the reconstructed frame $F_{v,t}$, the difference of the reconstructed frame $F_{v,t}$ and frame $F_{v,t+}$, and frame $F_{v,t+}$, respectively. The bottom row shows zoomed portions of the difference images in the top two rows. Errors were obtained by dropping one frame in V_1. The zoomed portions of the difference images show higher temporal consistency (represented by a smaller magnitude of the white color) of the our method compared to [13].

In regions with consistent texture, these blocks are sufficient to recover the lost blocks due to high spatial correlation between neighbouring blocks. But at the object boundaries, this correlation decreases and the spatially neighbouring top and left blocks are not as useful. This limitation is largely overcome by our approach, which can accurately track blocks at object boundaries in a neighbouring view. Compared to the method in [13], our method uses two frames from V_0 and V_2, respectively, while the method in [13] looks for matching pixels in a single reference view. Moreover, unlike the method

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

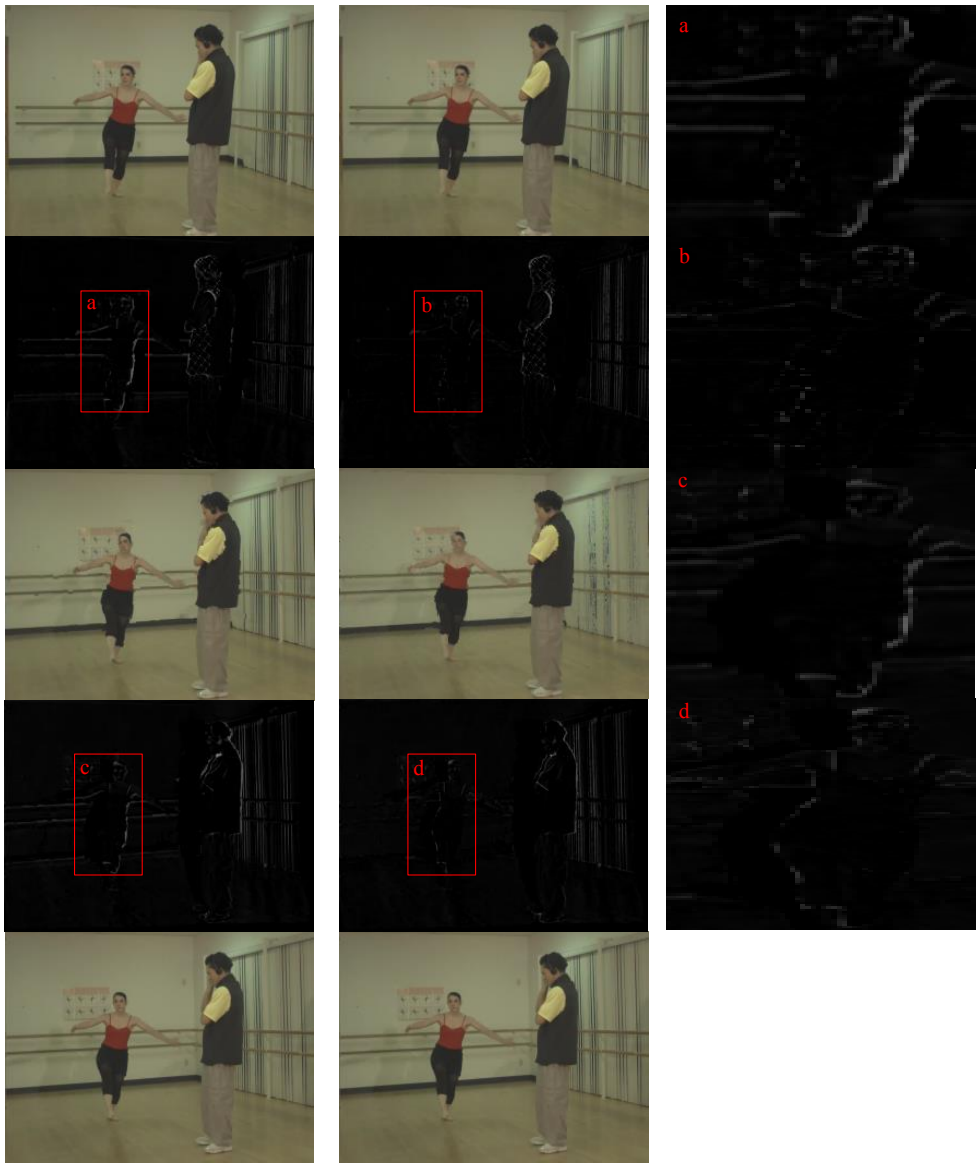


Fig. 8. Comparison of inter-view consistency for the Ballet sequence. Left: [13], Middle: our method, Right: Zoomed difference images. The first two columns show (from top to bottom): frame $F_{v^-,t}$, the difference of the warped frame from $F_{v,t}$ to $F_{v^-,t}$ and frame $F_{v^-,t}$, the reconstructed frame $F_{v,t}$, the difference of the warped frame from $F_{v,t}$ to $F_{v^+,t}$ and frame $F_{v^+,t}$, and frame $F_{v^+,t}$, respectively. The third column shows the zoomed difference images from the first two rows. Errors were obtained by dropping one frame in V_1. The zoomed parts of the difference images show higher inter-view consistency (smaller magnitude of the white color) of our method compared to [13].

in [13], our method checks different candidates before using one for concealment. The method of Liu et al. [12] uses motion information of a corresponding MB identified with the help of the global disparity between the current frame and a neighbouring view. It assumes a fixed disparity between two neighbouring frames from different views, which may not always be true. Our method gives better concealment results by including a candidate (MB_{r_4}) constructed with the help of camera parameters and depth information, which can accurately track pixels in a neighbouring view (see column P - P(w/o MB_r4) in the tables).

The increased gains for camera arrangements with short baseline distances can be attributed to the higher inter-view correlations in such settings. This does not only show that our method efficiently recovers the lost frames but that it also limits error propagation to other frames (see in particular Table III).

To analyse the time complexity of our method, we split it into two main steps:

- Step 1: for each macroblock of a lost frame, compute a set $\mathcal{C}$ of four candidate macroblocks ($MB_{r_1}, MB_{r_2}, MB_{r_3}$ and MB_{r_4}).
- Step 2: for each 4×4 sub-block of the macroblock, select the best 4×4 sub-block from the set $\mathcal{C}$ according to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

Fig. 9. Comparison of inter-view consistency for the Breakdancer sequence. Left: [13], Middle: proposed method, Right: Zoomed difference images. The first two columns show (from top to bottom): frame $F_{v-,t}$, the difference of the warped frame from $F_{v,t}$ to $F_{v-,t}$ and frame $F_{v-,t}$, the reconstructed frame $F_{v,t}$, the difference of the warped frame from $F_{v,t}$ to $F_{v+,t}$ and frame $F_{v+,t}$, and frame $F_{v+,t}$, respectively. The third column shows the zoomed difference images from the first two rows. Errors were obtained by dropping one frame in V_1. The zoomed parts of the difference images show higher inter-view consistency (smaller magnitude of the white color) of the proposed method compared to [13].

ICF metric.

In Step 1, only the computation of MB_{r_4} via view synthesis requires non-trivial operations. In Step 2, the selection process is very fast as only four candidates are considered. Similarly, the computation of the ICF metric, which is repeated 16 times for a given sub-block, is straightforward, with only the computation of positions (i_1, j_1) and (i_2, j_2) through the warping method requiring some efforts.

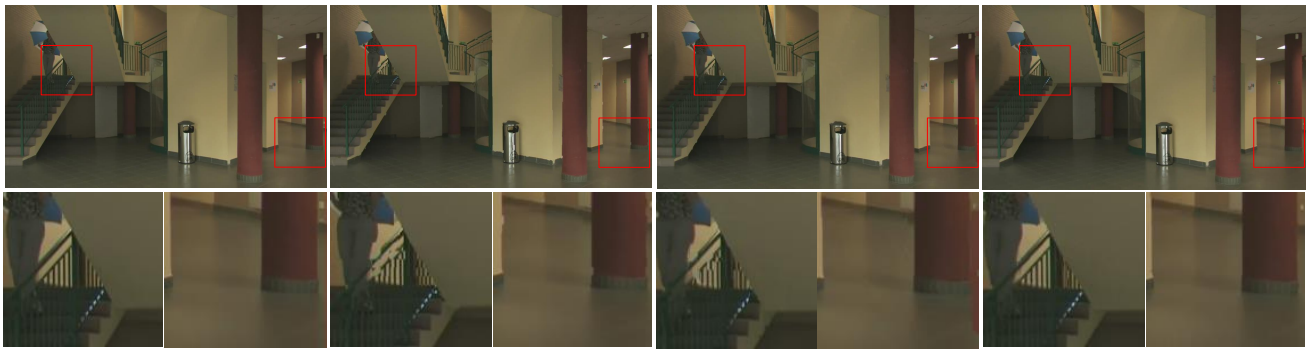
While our method is slower than the methods in [12] and [13], running time measurements indicate that it is suitable for broadcasting applications. For example, on a laptop with an Intel Core i5 Duo 2.67 GHz processor and 4 GB RAM, our method decodes the 100-frame Breakdancer sequence in 81.23 s compared with 52.08 s and 65.32 s for the methods in [12] and [13], respectively (for FLR = 5%).

V. Conclusion

We proposed a scene-consistent error concealment method to recover lost frames when compressed MVD video is broadcast over an error-prone delivery channel. Our method uses a cost function that combines temporal and view consistency criteria to reconstruct lost blocks from a set of candidate blocks. Simulation results show that our method does not only outperform conventional error concealment approaches in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

(a) 7th frame of Breakdancer sequence



(b) 4th frame of Poznan_Hall2 sequence

Fig. 10. Frame reconstruction. From left to right: no frame loss, reconstructed using [12], reconstructed using [13], and reconstructed using our method. For each sequence, the top row shows the full frame while the bottom row highlights parts of the frames. The lost frame is from V_1.

TABLE I
AVERAGE PSNR (dB) OVER CONCEALED FRAMES FOR DIFFERENT FRAME LOSS RATES (FLRS). P DENOTES OUR METHOD AND P(W/O MB_R4) DENOTES OUR METHOD WITHOUT VSC.

Sequence	FLR	No loss	P(w/o MB_r4)	BMC	[12]	[13]	P	P - P(w/o MB_r4)	P - BMC	P - [12]	P - [13]
Ballet	5%	37.05	26.10	26.50	25.89	26.72	27.08	0.98	0.58	1.19	0.36
	10%	36.97	25.20	25.34	25.16	25.63	26.67	1.47	1.33	1.51	1.04
	20%	37.02	24.69	24.73	24.78	25.24	26.36	1.67	1.63	1.58	1.12
Breakdancer	5%	35.90	27.33	26.79	27.44	27.13	27.47	0.14	0.68	0.03	0.56
	10%	35.80	26.24	26.39	26.71	26.29	26.77	0.53	0.38	0.06	0.48
	20%	35.79	25.73	26.04	26.31	25.58	26.34	0.61	0.40	0.03	0.76
Poznan_Street	5%	41.24	29.23	29.67	28.94	30.24	31.86	2.63	2.19	2.92	1.62
	10%	41.09	25.55	28.62	27.67	29.43	31.19	2.64	2.57	3.52	1.76
	20%	41.05	27.82	27.58	26.62	28.69	30.64	2.82	3.06	4.02	1.95
Poznan_Hall2	5%	44.26	34.39	34.82	34.68	35.05	36.45	1.46	1.63	1.77	1.40
	10%	44.07	34.11	33.94	33.81	34.28	35.76	1.65	1.82	1.95	1.48
	20%	44.04	33.17	33.29	33.09	33.70	35.40	2.23	2.11	2.31	1.70

reconstruction fidelity but also gives more consistent frames. The proposed consistent error concealment method can significantly improve the quality of MVD based 3D video that has been corrupted by transmission errors.

Our method is generic and flexible in the choice of the underlying error concealment methods that are used to generate candidate blocks. The choice of the methods to create candidate blocks for reconstruction in Section III-C is motivated by the idea that MBs reconstructed using view synthesis are expected to have better inter-view consistency while those obtained using motion compensation are expected to have better temporal consistency. Hence an appropriate selective combination of these methods based on an overall inconsistency evaluation criteria would result in frames that are consistent in both the inter-view and temporal directions. Another motivation is to make available a diverse set of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

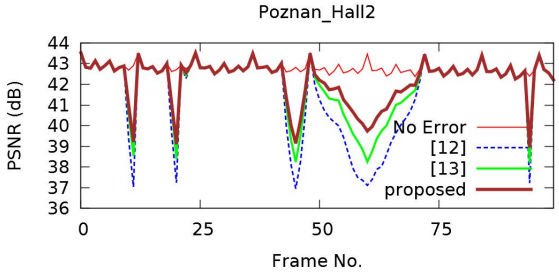
TABLE II
AVERAGE PSNR (dB) OVER ALL FRAMES FOR DIFFERENT FRAME LOSS RATES (FLRS). P DENOTES OUR METHOD AND P(W/O MB_R4) DENOTES OUR METHOD WITHOUT VSC.

Sequence	FLR	No loss	P(w/o MB_r4)	BMC	[12]	[13]	P	P - P(w/o MB_r4)	P - BMC	P - [12]	P - [13]
Ballet	5%	36.98	33.06	33.11	33.09	33.21	33.28	0.22	0.17	0.19	0.07
	10%	36.98	31.98	32.00	32.07	32.03	32.39	0.41	0.39	0.32	0.36
	20%	36.98	30.05	30.12	29.98	30.19	30.71	0.66	0.59	0.73	0.52
Breakdancer	5%	35.82	28.27	28.56	28.54	28.60	28.63	0.36	0.08	0.09	0.03
	10%	35.82	27.35	28.27	28.29	28.04	28.36	1.01	0.09	0.07	0.32
	20%	35.82	26.68	27.76	27.85	27.32	27.93	1.25	0.17	0.08	0.61
Poznan_Street	5%	41.17	36.75	37.16	36.61	37.01	37.41	0.66	0.35	0.80	0.40
	10%	41.17	35.26	35.95	34.87	35.80	36.64	1.38	0.69	1.77	0.84
	20%	41.17	33.53	33.98	31.39	33.93	35.10	1.57	1.12	3.71	1.17
Poznan_Hall2	5%	44.10	40.18	40.29	39.95	40.33	40.49	0.31	0.19	0.54	0.16
	10%	44.10	38.92	39.43	38.66	39.15	39.79	0.87	0.36	1.13	0.64
	20%	44.10	37.29	37.35	35.98	37.48	38.31	1.02	0.96	2.33	0.83

TABLE III
AVERAGE PSNR (dB) OVER CONCEALED FRAMES AND FRAMES DEPENDING ON THEM FOR DIFFERENT FRAME LOSS RATES (FLRS). P DENOTES OUR METHOD AND P(W/O MB_R4) DENOTES OUR METHOD WITHOUT VSC.

Sequence	FLR	No loss	P(w/o MB_r4)	BMC	[12]	[13]	P	P - P(w/o MB_r4)	P - BMC	P - [12]	P - [13]
Ballet	5%	36.96	27.96	28.11	27.54	28.42	29.28	1.32	1.17	1.74	0.86
	10%	36.96	26.89	27.00	26.87	27.96	28.39	1.50	1.39	1.52	0.97
	20%	36.97	26.58	26.72	26.91	27.49	28.71	2.13	1.99	1.80	1.22
Breakdancer	5%	35.79	27.10	26.98	26.54	27.28	27.63	0.53	0.65	1.09	0.35
	10%	35.77	26.23	26.27	25.89	26.49	26.99	0.76	0.72	1.10	0.50
	20%	35.81	25.83	25.76	25.31	25.89	26.58	0.75	0.82	1.27	0.69
Poznan_Street	5%	41.15	30.03	30.29	29.98	30.66	32.41	2.38	2.12	2.43	1.75
	10%	41.13	29.02	29.15	29.05	29.61	31.64	2.62	2.49	2.59	2.03
	20%	41.16	28.35	27.41	27.12	28.84	31.10	2.75	3.69	3.98	2.26
Poznan_Hall2	5%	44.08	35.04	35.53	35.20	35.95	37.49	2.45	1.96	2.29	1.54
	10%	44.09	34.01	34.60	34.32	35.27	36.79	2.78	2.19	2.47	1.52
	20%	44.07	32.85	33.95	33.53	34.66	36.31	3.46	2.36	2.78	1.65

(a)



(b)

Fig. 11. PSNR vs. frame number at 5% FLR for (a) the Poznan_Street and (b) the Poznan_Hall2 sequences.

candidate blocks such that the concealment process is not dependent on the availability of a particular frame.

In our simulations, the value of the weighting factor α in (5) gives the same importance to temporal and inter-view inconsistencies, which is not necessarily the best choice. Similarly, using the same α for all 4×4 blocks gives the same importance to blocks with occluded regions as to those without. Adapting α according to the scene or requirements may lead to better results and is left as future work. For example, our method could be extended to detect occluded pixels in the 4×4 blocks and use a smaller weighting factor for blocks with fewer occluded pixels.

Another direction for future work is to consider cases where whole-frame loss is not assumed. In this context, applying the proposed approach in conjunction with spatial error concealment techniques (e.g., [11] and [24]) for recovering block losses may prove to be very effective.

ACKNOWLEDGMENT

We thank Poznan University of Technology for providing the "Poznan_Hall2" and "Poznan_Street" test sequences.

REFERENCES

[1] R. Di Bari, M. Bard, A. Arrinda, P. Ditto, G. Araniti, J. Cosmas, K. K. Loo, and R. Nilavalan, "Measurement campaign on transmit delay diversity for mobile DVB-T/H systems, *IEEE Trans. Broadcast.*, vol. 56, no. 3, pp. 369-378, Sep. 2010

[2] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP'07)*, Texas, USA, Sept. 2007.

[3] F. Shao, G. Jiang, M. Yu, K. Chen, and Y. Ho, "Asymmetric coding of multi-view video plus depth based 3-D video for view rendering," *IEEE Transactions on Multimedia*, vol. 14, no. 1, pp. 157–167, Feb. 2012.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

[4] J. Lee, H. Wey, and D. Park, "A fast and efficient multi-view depth image coding method based on temporal and inter-view correlations of texture images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 12, pp. 1859–1868, Dec. 2011.

[5] M. Hannuksela, D. Rusanovskyy, S. Wenyi, L. Chen, R. Li, P. Aflaki, D. Lan, M. Joachimiak, H. Li, and M. Gabbouj, "Multiview-video-plus-depth coding based on the advanced video coding standard," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3449–3458, Sept. 2013.

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

[7] C. Hellge, T. Schierl, and T. Wiegand, "Mobile TV using scalable video coding and layer-aware forward error correction," *Proc. IEEE Int. Conf. Multimedia Expo*, Hannover, Germany, June 2008, pp. 1177–1180.

[8] W. Fischer, "Digital Video and Audio Broadcasting Technology: A Practical Engineering Guide," Springer, 2010.

[9] K. D. Singh, G. Rubino, "Quality of Experience estimation using frame loss pattern and video encoding characteristics in DVB-H networks," in *Proc. 18th International Packet Video Workshop (PV2010)*, Hong Kong, China, pp. 150–157, 13-14 Dec. 2010.

[10] K. Song, T. Chung, Y. Oh, and C. Kim, "Error concealment of multi-view video sequences using inter-view and intra-view correlations", *J. Visual Commun. Image Represent.*, vol. 20, no. 4, pp. 281–292, May 2009.

[11] Y. Liu, J. Wang, and H. Zhang, "Depth image based temporal error concealment for 3-D video transmission," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 600–604, April 2010.

[12] S. Liu, Y. Chen, Y. Wang, M. Gabbouj, M. Hannuksela, and H. Li, "Frame loss error concealment for multiview video coding," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS2008)*, pp. 3470–3473, 18-21 May 2008.

[13] T. Chung, S. Sull, and C. Kim, "Frame loss concealment for stereoscopic video plus depth sequences," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1336–1344, August 2011.

[14] C. Hewage, S. Warrall. S. Dogan, and A. Kondoz, "Frame concealment algorithm for stereoscopic video using motion vector sharing," in *Proc. ICME 2008*, Germany, June 2008.

[15] B. Yan and J. Zhou, "Efficient frame concealment for depth image based 3-D video transmission," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 941–945, June 2012.

[16] W. Lam, A. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Minneapolis, MN, Apr. 1993, pp. 417–420.

[17] JMVC 8.5, garcon.ient.rwthaachen.de, 2011.

[18] T. Wiegand, G. Sullivan G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[19] Y. Chen, K. Yu, J. Li, and S. Li, "An error concealment algorithm for entire frame loss in video transmission," in *Proc. IEEE Picture Coding Symp.*, 2004, San Francisco, CA.

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

[21] X. Ji, D. Zhao, and W. Gao, "Concealment of whole-picture loss in hierarchical B-picture scalable video coding," *IEEE Transactions on Multimedia*, vol. 11, pp. 11–22, Jan. 2009.

[22] Y. Guo, Y. Chen, Y. Wang, H. Li, M. Hannuksela, M. Gabbouj, "Error resilient coding and error concealment in scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 6, pp. 781–795, June 2009.

[23] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[24] W. Kung, C. Kim, and C. Kuo, "Spatial and temporal error concealment techniques for video transmission over noisy channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 789–803, 2006.

[25] T. Basha, Y. Moses, and S. Avidan, "Geometrically consistent stereo seam carving," in *Proc. IEEE International Conference on Computer Vision (ICCV2011)*, pp. 1816–1823, 6-13 Nov. 2011.

[26] G. Floros and B. Leibe, "Joint 2d-3d temporally consistent semantic segmentation of street scenes," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2012)*, pp. 2823–2830, 2012.

[27] A. Tankus and Y. Yeshurun, "Scene-consistent detection of feature points in video sequences," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 1–29, 2005.

[28] T. Maugey, P. Frossard, and G. Cheung, "Consistent view synthesis in interactive multiview imaging," in *Proc. IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Oct. 2012, pp. 2717–2720.

[29] I. Ahn and C. Kim, "Depth-based disocclusion filling for virtual view synthesis," in *Proc. ICME2012*, 2012, pp. 109–114.

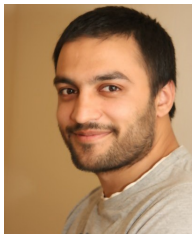
[30] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View synthesis for multiview video compression," in *Proc. Picture Coding Symposium (PCS06)*, Beijing, China, Apr. 2006.

[31] R. Tsai, "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[32] H. Koo, Y. Jeon, and B. Jeon, "MVC motion skip mode," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-W081.

[33] L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics*, vol. 23, no. 3, Aug. 2004.

[34] M. Domanski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner "Poznan Multiview Video Test Sequences and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, China, Oct. 2009.



Shadan Khattak received the BSc degree in computer systems engineering from University of Engineering and Technology (UET), Peshawar, Pakistan, in 2007, the MSc degree in data communications from The University of Sheffield, Sheffield, UK, in 2009, and the PhD degree in multimedia communication from De Montfort University, Leicester, UK, in 2014. From 2007 to 2008, and 2009 to 2010, he was a Lecturer in the department of electrical engineering at COMSATS Institute of Information Technology, Abbottabad, Pakistan and from 2013 to 2014, he was a Research Engineer at Parabola Research Ltd., Southampton, United Kingdom. He is currently working as an Assistant Professor in the department of electrical engineering at COMSATS Institute of Information Technology, Abbottabad, Pakistan. His research interests include 2D and 3D video compression, fast video encoding algorithms, and error concealment techniques

Thomas Maugey (S'09, M'11) graduated from École Supérieure d'Électricité, Supélec, Gif-sur-Yvette, France in 2007. He received the MSc degree in fundamental and applied mathematics from Suplec and Universit Paul Verlaine, Metz, France, in 2007. He received his PhD degree in Image and Signal Processing at TELECOM ParisTech, Paris, France in 2010. From October 2010 until September 2014, he was a postdoctoral researcher at the Signal Processing Laboratory (LTS4) of Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. Since November 2015, he has been a research scientist at INRIA in the team-project SIROCCO, in Rennes, France. His research interests include monoview and multiview distributed video coding, 3D video communication, data representation, video compression, network coding and view synthesis.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2418631, IEEE Transactions on Circuits and Systems for Video Technology

Raouf Hamzaoui (M'02, SM'07) received the MSc degree in mathematics from the University of Montreal, Canada, in 1993, the Dr.rer.nat. degree from the University of Freiburg, Germany, in 1997 and the Habilitation degree in computer science from the University of Konstanz, Germany, in 2004. He was an Assistant Professor with the Department of Computer Science of the University of Leipzig, Germany and with the Department of Computer and Information Science of the University of Konstanz. In September 2006, he joined De Montfort University where he is a Professor in Media Technology and Head of Research and Innovation for the Faculty of Technology. His research interests include image and video coding, multimedia communication systems, channel coding, and error control systems.

Shakeel Ahmad received the PhD (Dr.-Ing) degree from the University of Konstanz, Konstanz, Germany, in 2008, the MSc degree in Information and Communication Systems from Hamburg University of Technology, Hamburg, Germany, in 2005 and the BSc (Hons) degree in Electronics and Communication Engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2000. He is currently a Senior Research Fellow in the Faculty of Technology at De Montfort University, UK. His research interests include video coding and streaming, channel coding, P2P networks, communication protocols, and MANETs.

Pascal Frossard (S'96, M'01, SM'04) received the M.S. and PhD degrees, both in electrical engineering, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997 and 2000, respectively. Between 2001 and 2003, he was a member of the research staff at the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he worked on media coding and streaming technologies. Since 2003, he has been a faculty at EPFL, where he heads the Signal Processing Laboratory (LTS4). His research interests include image representation and coding, visual information analysis, distributed image processing and communications, and media streaming systems.