Frame Loss Error Concealment For Multiview Video Coding

Shujie Liu[1], Ying Chen[2], Ye-Kui Wang[3], Moncef Gabbouj[2], Miska M. Hannuksela[3], Houqiang Li[1]

[1]University of Science and Technology of China
Hefei, China
{lsjxbd@mail., lihq@}ustc.edu.cn

[2]Institute of Signal Processing
Tampere University of Technology
Tampere Finland
{ying.chen, moncef.gabbouj}@tut.fi

[3]Nokia Research Center
Tampere Finland
ye-kui.wang@nokia.com
miska.hannuksela@nokia.com

***Abstract*—The Multiview Video Coding (MVC) standard is currently under development by the Joint Video Team as an extension of the Advanced Video Coding (H.264/AVC) standard. An MVC encoder compresses more than one viewpoint of a scene captured by different cameras. Redundancies between views can be used for inter-view prediction in encoding as well as error concealment in decoding. In this paper, a new algorithm utilizing motion information of pictures from other views to conceal a lost picture is proposed. The algorithm first derives motion information for a lost picture based on motion fields of pictures in adjacent views. Then, traditional motion compensation is invoked within the view containing the lost picture to derive a concealed frame. Experimental results show that the proposed algorithm can improve video quality with a negligible computational complexity overhead compared to simple temporal error concealment algorithms.**

I. INTRODUCTION

Multiview video technologies have gained significant interest recently. In multiview video coding (MVC), the original video content is a group of video sequences captured by multiple cameras at the same time for the same scene but from different viewpoints. Typical multiview applications include free-viewpoint video, where the viewer can interactively choose his/her viewpoint in a three-dimensional (3D) space to observe a scene from a preferred perspective [1], and 3D television (TV), where multiple views are displayed simultaneously [2]. Due to the huge amount of original video data, the transmission part of multiview video systems relies heavily on compression of the video captured by multiple cameras. Fortunately, sophisticated coding tools can utilize correlations among different views to compress multiview content better than coding each view independently. One of the goals for the development of the MVC standard [3] is to discover such tools. Moreover, redundancies between views also provide opportunities for improved error concealment.

During transmission of video data, packet losses may occur. This may lead to undesirable effects such as system instability, unacceptable video quality and unpredictable decoder behavior. Therefore, techniques to control the impacts of transmission errors are highly desirable. A number of error control algorithms operating in the source coding layer, often referred to as error concealment methods, have been proposed in the literature [4][5]. They can be summarized into three categories: encoder and decoder interactive error concealment, error resilient encoding, and decoder error concealment [6]. Interactive error concealment typically utilizes the feedback from the receiver to adjust the encoding strategy to minimize error propagation. Error resilient encoding mainly utilizes some redundant information added at the encoder side. In this paper, we mainly focus on decoder error concealment. Some algorithms based on decoder error concealment have been proposed in [7][8][9]. They make use of temporal or spatial correlation between the macroblocks (MBs) in damaged area and its adjacent MBs in the same or previous frame. These algorithms assume that if either a single MB or a slice consisting of several consecutive MBs is lost, information from the neighboring available MBs or MBs in the adjacent frames can be used to estimate both motion vectors and texture of the missing MB. However, in some applications, a coded picture typically fits in one packet, and a transmission error will lead to a loss of a whole slice or frame.

Algorithms handling frame loss have also been proposed. For example, Belfiore et al. [10] addressed a pixel-level algorithm based on the optical flow theory, assuming that the motion between two consecutive pictures does not vary in a dramatic way. It exploited motion information in a few past frames to estimate forward motion of a lost frame. An error concealment algorithm called "BLSkip" was introduced for scalable video coding [11]. This algorithm gets motion vectors for the lost frame from the co-located MBs of the lower layers and outperforms simple temporal error concealment methods, such as "Frame Copy", where a lost frame is reconstructed by copying the reconstruction of the previous frame.

Getting motion information for a lost frame is important for the performance of an error concealment algorithm. In an MVC bitstream, there is usually a high correlation between two views. This correlation, such as motion field similarity, can be used to get a more accurate estimation of coding modes and motion vectors. In this paper, an algorithm utilizing this correlation is proposed. Experimental results show that the proposed algorithm can improve video quality comparing to low complexity temporal error concealment algorithms.

II. BACKGROUND

In this section, a brief overview of MVC is included first. Then, two error concealment methods, which are used for comparison, are discussed separately: (1) Frame Copy (FC); (2) Temporal Direct motion vector generation (TD). FC method can be used to conceal a lost picture of any type, while TD method conceals lost B pictures.

978-1-4244-1684-4/08/$25.00 ©2008 IEEE

A. Overview of MVC

An H.264/AVC bitstream consists of NAL units. NAL units can be categorized into video coding layer (VCL) NAL units and non-VCL NAL units. Coded pictures are contained in VCL NAL units. MVC, as an extension of H.264/AVC, has the same concepts for NAL and VCL.

The current coding scheme of MVC has a structure shown in Fig. 1. For each view, a prediction structure with hierarchical B pictures is used. There is a base view which is H.264/AVC compliant and independently coded. A view V can rely on other views for decoding. For view V, they are dependent views. Typically each view with an even view identifier (called even view for simplicity) has the previous even view as the dependent view and an odd view has the previous and next even views as dependent views. Normally, each GOP (Group of Pictures) contains an anchor picture and the rest pictures in the GOP are non-anchor pictures. An anchor picture is a picture that can be correctly decoded without the decoding of previous pictures in decoding order. Anchor pictures can serve as random access points.

B. Error concealment of frame copy (FC)

Frame copy is a simple error concealment method. In this algorithm, a lost picture is handled as P picture and a reference picture list construction process (with or without reference picture list reordering) is invoked to construct a reference list in a way that the first reference picture in the list is the previous picture of this lost picture in output order . Then, each pixel value of the lost picture is copied from the corresponding pixel of that first reference picture.

C. Temporal direct motion vector generation (TD)

TD is also an error concealment method that usually uses information in the same view [12]. In this algorithm, for each lost MB, the motion vectors (MVs) and references are generated as if it is coded using the "temporal direct mode", which is illustrated in Fig. 2.

As shown in Fig. 2, the motion vectors of the direct-mode are found in the first picture of Reference Picture List 1 (RefPicList1). The motion vector of the co-located block MV_C is scaled to obtain MV_0 and MV_1 based on the temporal distance TD_B and TD_D. Furthermore, in simulation, FC method is used as alternation for lost P pictures and I pictures.

III. Proposed Algorithm

In this section, the relationship between motion fields of different views is discussed first based on a geometric model. After that, the proposed algorithm, called Motion Prediction (MP), is described.

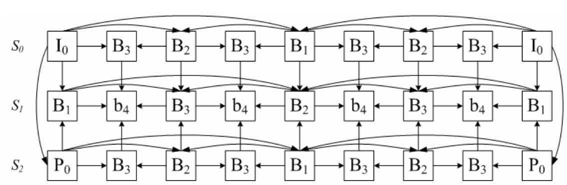


Figure 1. Inter-view temporal prediction structure

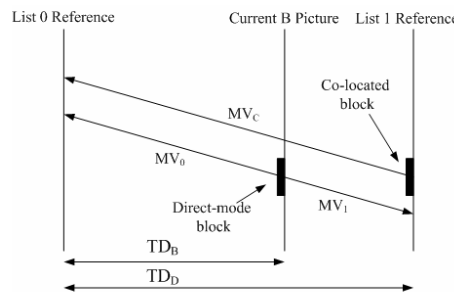


Figure 2. Example of temporal direct mode (MV generation)

A. Similarity of Motion Fields

Without loss of generality, we take the case of two-view video content as an example. Let P_t^0 denote the location of one pixel in view 0 at the time of t, while P_t^1 denote the location of the corresponding pixel in view 1. Moreover, P_{t-1}^i is the corresponding pixel of P_t^i at the time of t-1 shifted according to the temporal motion vectors MV_t^i. We can obtain the following equation:

$$P_t^i = P_{t-1}^i + MV_t^i \quad (1)$$

Normally, a six-parameter affine model is used to describe the global disparity between two views. The model includes a 2x2 transformation matrix $\boldsymbol{A}$ and a 2x1 displacement matrix $\boldsymbol{B}$. $\boldsymbol{A_t}$ and $\boldsymbol{B_t}$ are supposed to represent the affine model between two views at the time of t.

In this paper, the global disparity between two views is simplified to a two parameter model, which contains only the displacement between the two views, and the matrix $\boldsymbol{A}$ is simplified to an identity matrix. Furthermore, the geometrical locations and angles of the cameras are assumed to be identical during a short interval between a picture and its reference picture. Combining this assumption and (1), we can obtain:

$$MV_t^1 = MV_t^0 \quad (2)$$

If extended to a block, this equation means that the motion fields of corresponding blocks in two views do have similarity. And this similarity is used in the proposed algorithm, which is described in the following part.

B. Algorithm Description

In the encoder side, the global disparity motion regarding the inter-view reference picture relative to the base view is sent for every anchor picture. In the decoder, the MP algorithm is described as follows.

When a picture is detected as lost, each MB of the lost frame is processed as follows. First, the corresponding MB (CMB) in a dependent view is found according to the global disparity motion. As indicated by (2), the mode and motion vectors of CMB are then copied for the lost MB. Finally, motion compensation is used to generate the concealed picture. There are cases when the forward method is not applicable for some special MBs or even a whole slice. They are concealed by other methods in MP algorithm:

If the CMB does not contain motion information, e.g., it is Intra coded, spatial error concealment for this MB is utilized. If the slice is a P slice, the current MB is set to skip mode in P picture; if the slice is a B slice, spatial direct mode is utilized.

If an anchor picture in one view that is not the base view is lost (it is detected to be an anchor picture if its corresponding picture in the base view with the same time instance is an anchor picture), it is concealed by copying sample values from that corresponding anchor picture of its dependent view.

IV. SIMULATIONS

A. Simulation conditions

In our simulations, the packet loss model in [13] was used, and we only focused on the cases of two views: view 0 and view 1, where view 0 is the dependent view of view 1. Two simulation models are used: simulation 1 and simulation 2. In simulation 1, each NAL unit was assumed to be contained in one packet. While in the other, one NAL unit was segmented into several 1400 bytes packets, if one of the segmentations is lost, the NAL unit is thought lost, and anchor pictures are assumed to be error free.

Only entire frame losses were considered, i.e., one slice per frame was encoded, which could be simply extended to slice loss cases. Different loss conditions were used in the simulation, with lower loss rate for view 0 than for view 1, which are all listed in TABLE I. The video sequences for the MVC common test conditions, *Akko&Kayo*, *Ballroom*, *Breakdancers*, *Exit*, *Race1*, *Rena*, and *Flamenco2*, were used but with only two views. Our implementation was based on the reference software of JMVM with a version of 4, and the common encoding settings were followed [14].

B. Results

The proposed error concealment method, FC, and TD were all tested. The performance comparison of the former two methods is reported in TABLE I. Fig. 3 shows the results of the three methods for the *Akko&Kayo, Exit* and *Rena* sequences at some conditions with no loss of view 0. It is clear that the proposed algorithm has a better performance than FC and TD both at the low packet loss rate and the high packet loss rate for *Akko&Kayo*, and has a little loss compared to TD for *Exit* in some conditions. Decoded pictures with degradation because of error propagation are shown in Fig. 4. They are the 162^{nd} frame of *race1* and correspond to the three different error concealment algorithms. As shown in Fig.4, the proposed algorithm introduces less degradation.

The experimental results show that the proposed algorithm obtained significant gains compared to the FC and TD error concealment algorithms, up to more than 2.6 dB and on average about 0.97 dB compared with FC algorithm in terms of average luma peak signal-to-noise ratio (PSNR).

C. Discussion

There are cases when MP is close to FC or even a little worse than TD, mainly because most of the motion is local and the estimated disparity is not accurate enough for every MB. For simplicity, we used a simple displacement to model the transformation between two views. However, in practice, the optimal transformation between two views maybe non-linear and objects with different depths and different location may need different disparities. Those effects, however, can be compensated by local disparity motion vectors, e.g. in MB level. A potential improvement can use algorithms to estimate disparity motion vectors for each MB. Based on the assumption that the disparity remains the same during a short interval, we can estimate the disparity of each MB according to the adjacent correct decoded pictures in other views.

In the case discussed above, temporal motion field similarity can be better extracted and can predict the motion field well for the lost frame. That is why TD has close performance as MP, e.g. for *Exit*. Another improvement can adaptively choose the best motion prediction from either temporal or view dimension.

V. ACKNOWLEDGMENT

The work of Shujie Liu and Houqiang Li are partially supported by NSFC General Program under contract No. 60572067 and NSFC Key Program under contract No. 60736043.

VI. CONCLUSION

As multiview technologies become more and more mature, it is necessary to consider error concealment algorithms for coded multiview bitstreams. Error concealment can take advantage of inter-view redundancies. In this paper, a fast error concealment algorithm was proposed for multiview video coding. Motion field similarity between different views was utilized to generate the motion vectors for each MB of a lost picture, by predicting them from the motion vectors of the corresponding MB in another view. The corresponding MB can be estimated by several means, while in this paper we used the global motion disparity. The proposed method can conceal entire missing frame of a video sequence with low complexity and high quality. Simulation results showed that this method has significant gain over the frame copy method.

REFERENCES

[1] "Requirements on Multi-view Video Coding v.5," document N7539, MPEG, Nice, France, Oct. 2005.

[2] A. Vetro, W. Matusik, H. Pfister, J. Xin, "Coding Approaches for End-to-End 3D TV Systems," Picture Coding Symposium, 2004.

[3] "Joint Draft 4.0 on Multiview Video Coding", JVT-X209, Geneva, Swizerland, Jun.-Jul. 2007.

[4] B. W. Wah, X. Su, and D.Lin, "A survey of error-concealment schemes for real-time audio and video transmissions over the internet," Proc. Int. Symp. Mutimedia Software Engineering, Dec. 2000.

[5] P. Cuenca, L. Orozco-Barbosa, A. Garrido, F. Quiles, and T. Olivares, "A survey of error concealment schemes for MPEG-2 video communications over ATM networks," Proc. IEEE 1997 Can. Conf. Elect.and Comput, Eng., May 1997.

[6] Y. Wang, S. Wenger, J. Wen, and A.K. Katsaggelos, "Error resilient video coding techniques," IEEE Signal Processing Magazine, vol. 17, issue 4, pp. 61 - 82, Jul. 2000.

[7] J. Zheng and L. P. Chau, "A temporal error concealment algorithm for H.264 using Lagrange interpolation," International Symposium on Circuits and Systems, ISCAS 2004

[8] L. W. Kang and J. J. Leou, "A hybrid error concealment scheme for MPEG-2 video transmission based on best neighborhood matching algorithm," J. Visual Communication and Image Representation, vol. 16, issue 3, pp.288-310, 2005.

[9] Y-K. Wang, M. M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, "The error concealment feature in the H.26L test model," ICIP 2002, Sep, 2002.

[10] S. Belfiore, M. Grangetto, E. Magli and G. Olmo, "Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flow estimation," IEEE Trans. on Multimedia, vol 7, issue 2, pp. 316-329, Apr. 2005.

[11] Y. Chen, K. Xie, F. Zhang, P. Pandit, J. Boyce, "Frame Loss Error Concealment for SVC", Journal of Zhejiang University SCIENCE A, also in Proc. Packet Video Apr. 2006.

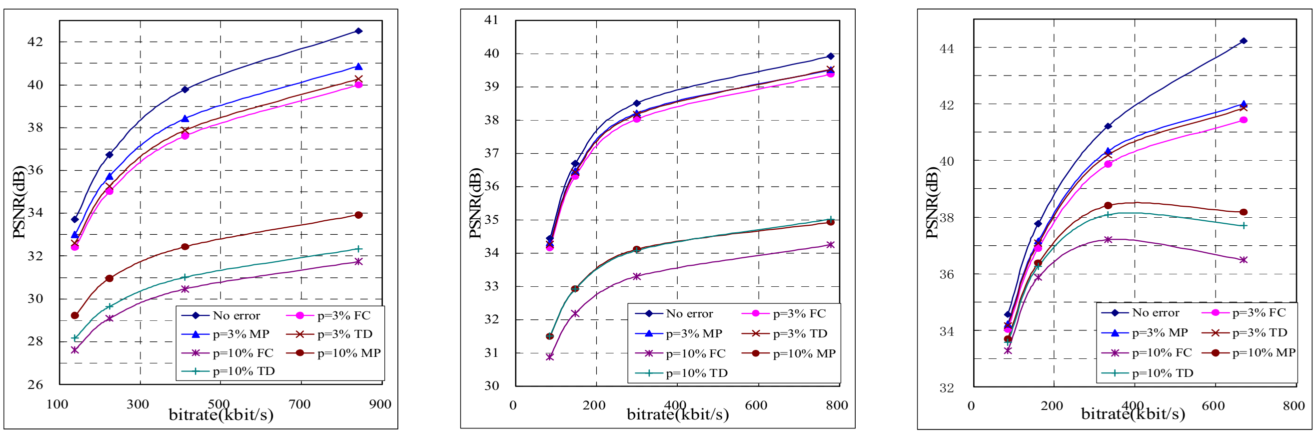
[12] ITU-T Rec. H.264 | ISO/IEC 14496-10, "Advanced video coding for generic audiovisual services, v3: 2005".

[13] S. Wenger, "Error Patterns for Internet Experiments," VCEG Q15-I-16r1, 2002.

[14] "Common Test Conditions for Multiview Video Coding," JVT-T207, Klagenfurt, Austria, Jul. 2006.

TABLE I. PSNR BETWEEN MP AND FC WHEN DIFFERENT PACKET LOSS RATES FOR VIEW 0 (Q) AND VIEW 1 (P) ARE APPLIED

Sequence	△PSNR(dB) (MP vs FC) (Simulation 1)						△PSNR(dB) (MP vs FC) (Simulation 2)					
	q = 0% p = 3%	*q = 0% p = 5%*	*q = 0% p = 10%*	*q = 3% p = 5%*	*q = 3% p = 10%*	*q = 5% p = 10%*	*q = 0% p = 3%*	*q = 0% p = 5%*	*q = 0% p = 10%*	*q = 3% p = 5%*	*q = 3% p = 10%*	*q = 5% p = 10%*
Akko&Kayo	0.77	1.15	1.93	1.16	1.43	1.16	1.20	1.91	2.63	1.53	2.13	1.47
ballroom	0.12	0.71	0.97	0.77	0.93	0.92	0.50	1.29	1.57	1.09	1.44	1.26
Breakdancers	0.14	0.56	0.53	0.66	0.56	0.67	0.47	1.04	1.01	1.05	1.06	1.11
exit	0.17	0.59	0.76	0.53	0.75	0.74	0.36	0.64	0.61	0.63	0.63	0.59
Flamenco2	0.24	0.31	0.48	0.44	0.62	0.62	0.40	0.79	1.07	0.92	1.18	1.15
race1	0.67	1.33	2.30	1.17	1.51	0.93	1.46	2.23	2.65	1.36	1.98	1.55
rena	0.34	0.45	0.76	0.57	0.76	0.71	0.37	0.67	0.93	0.61	0.96	0.89
(Average)	0.35	0.73	1.10	0.76	0.94	0.82	0.68	1.22	1.50	1.03	1.34	1.15



(a) Akko&Kayo p=3%, 10% (Simulation 1) (b) Exit p=3%, 10% (Simulation 1) (c) Rena p=3%, 10% (Simulation 2)

Figure 3. Comparison MP, TD, FC for Akko&Kayo, Rena and Exit sequences with different packet loss rates (p) in view 1

(a) Cropped picture corresponds to FC (b) Cropped picture corresponds to TD (c) Cropped picture corresponds to MP

Figure 4 Subjective quality comparison of pictures for "race1", with no loss in view 0 and 10%, loss rate in view 1, QP = 37.