AUTO-REGRESSIVE MODEL BASED ERROR CONCEALMENT SCHEME FOR STEREOSCOPIC VIDEO CODING

Xinguang Xiang[1], Debin Zhao[1], Siwei Ma[2], and Wen Gao[2]

[1] School of Computer Science and Technology, Harbin Institute of Technology,
Harbin 150001, P.R. China
[2] Institute of Digital Media, Peking University, Beijing, 100871, P.R. China

ABSTRACT

Stereoscopic video is an important manner for 3-D video applications, and robust stereoscopic video transmission has posed a technical challenge for stereoscopic video coding. In this paper, an auto-regressive (AR) model based error concealment scheme is proposed for stereoscopic video coding to address the challenging problem. The proposed error concealment scheme includes a temporal AR model for independent view, and a temporal-interview AR model for inter-view predicted view. First, appropriate motions and disparities for lost blocks are derived. Then, the proposed AR model coefficients are computed according to the spatial neighboring pixels and their temporal-correlated and interview-correlated pixels indicated by the selected prediction directions. Finally, applying the AR model, each pixel of the lost block is interpolated as a weighted summation of pixels in the reference frame along the selected prediction directions. Simulation results show that the performance of the proposed scheme is superior to conventional temporal error concealment methods for stereoscopic video coding.

Index Terms— Stereoscopic video coding, error concealment, auto-regressive model, temporal, temporal-interview

1. INTRODUCTION

Recent years have seen increased research efforts for stereoscopic video coding, as stereoscopic video has a wide variety of 3-D applications. Stereoscopic video, which consists of video sequences of the same scene captured by two cameras from different views at the same time, is an important form of 3-D video. Different from single-view video, stereoscopic video often has strong inter-view correlations among different views. Thus the popular stereoscopic video coding frameworks often adopt both the disparity compensation prediction (DCP) [1] for the inter-view prediction and motion compensation prediction (MCP) for the single-view prediction.

On the other hand, robust video transmission over error-prone networks is a technical challenge for video coding. Error concealment, which is performed at the decoder side to fill up the lost video contents, is a major technique to deal with the challenge. Error concealment in the traditional single view coding has been widely exploited [2-6]. Temporal replacement (TR) [2], which utilizes the zero motion vector (MV) to reconstruct a lost macroblock (MB), is a simpler error concealment scheme. A very popular and widely accepted technique, the boundary matching algorithm (BMA), is proposed in [3-4] to select an optimal MV to substitute for the lost one. In [5], Chen et al. proposed a technique which combines the overlapped motion compensation and the side match criterion. Refined boundary matching algorithm (RBMA) is proposed to conceal different regions of a lost block with different motion vectors in [6]. However, these techniques are all designed for single-view video coding, so they will be inefficient if directly applied to stereoscopic video coding, as the inter-view correlation is not considered and exploited.

To our knowledge, few works [7] have been reported on the error concealment for stereoscopic video coding. How to effectively perform error concealment for stereoscopic video coding is still an unanswered question. Auto-regressive (AR) model [8], which is an efficient description of random process, is able to have desirable performance for interpolation. In this paper, utilizing the temporal and interview correlation in stereoscopic video and the superior property of the AR model, we propose an auto-regressive (AR) model based error concealment scheme for stereoscopic video coding. The proposed error concealment scheme includes a temporal AR model for independent view, and a temporal-interview AR model for inter-view predicted view. First, the boundary matching algorithm (BMA) is utilized to select a best motion vector (MV) and a best disparity vector (DV) for each lost block from the prediction vectors of available neighboring blocks. Then, we compute our proposed AR model coefficients according to the spatial neighboring pixels and their temporal-correlated and interview-correlated pixels indicated by the selected best prediction directions. Finally, applying the AR model, each pixel of the lost block is interpolated as a weighted

summation of pixels in the reference frame along the selected best prediction directions.

The rest of this paper is organized as follows. Section 2 describes the proposed algorithm in detail. Section 3 reports the simulation results. Section 4 concludes this paper.

2. THE PROPOSED AUTO-REGRESSIVE MODEL BASED ERROR CONCEALMENT

The basic prediction structure of stereoscopic video coding is shown in Fig. 1, where the left view is predicted only by MCP, called independent view, and the right view is predicted by both MCP and DCP, called inter-view predicted view.
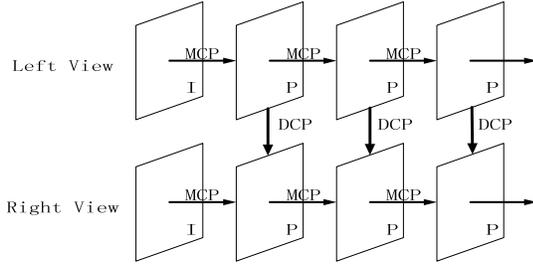

Fig. 1 The basic prediction structure of the two-view based stereoscopic video coding

The proposed AR model aims to recover the lost data in current frame, based on the picture data in correlated frames from temporal and interview directions indicated by appropriate motions and disparities. Fig.2 illustrates the proposed AR model based error concealment scheme. We utilize the BMA technique to select appropriate prediction directions for the lost MB, and utilize the pixels in neighboring MBs to derive AR model coefficients. Finally, we recover the lost pixels by their temporal-correlated and interview-correlated pixels with the proposed AR model. Note that the proposed scheme is composed of a temporal AR model for independent view, and a temporal-interview AR model for inter-view predicted view.

2.1. Selection of prediction directions

In order to choose appropriate prediction vectors which are used as the prediction directions for AR model, we utilize the BMA criterion. The cost function of BMA is defined as the absolute difference between the external boundary of the lost MB in the current frame and the internal boundary of the replacing MB in the reference frame, and it is formulated as follows:

$$Cost_{BM} =$$
$$\sum_{x=x_0}^{x=x_0+15} [|P(x,y_0-1)-P^r(x+v_x,y_0+v_y)|+|P(x,y_0+16)-P^r(x+v_x,y_0+15+v_y)|]$$
$$+ \sum_{y=y_0}^{y=y_0+15} [|P(x_0-1,y)-P^r(x_0+v_x,y+v_y)|+|P(x_0+16,y)-P^r(x_0+15+v_x,y+v_y)|]$$

(1)

where (x_0, y_0) denotes the coordinate of the top-left pixel in the lost MB, and (v_x, v_y) denotes the candidate vector. P and P^r denote the pixels of the current and reference frames, respectively. For each lost MB, we can get some candidate motion vectors and disparity vectors from neighboring available prediction vectors. While the zero motion vector and the zero disparity vector are also candidates. The motion vector which results in the smallest temporal direction cost is selected as the best motion. Similarly the disparity vector which results the smallest interview direction cost is selected as the best disparity. Note that, only motions are selected for independent view while both motions and disparities are selected for inter-view predicted view.
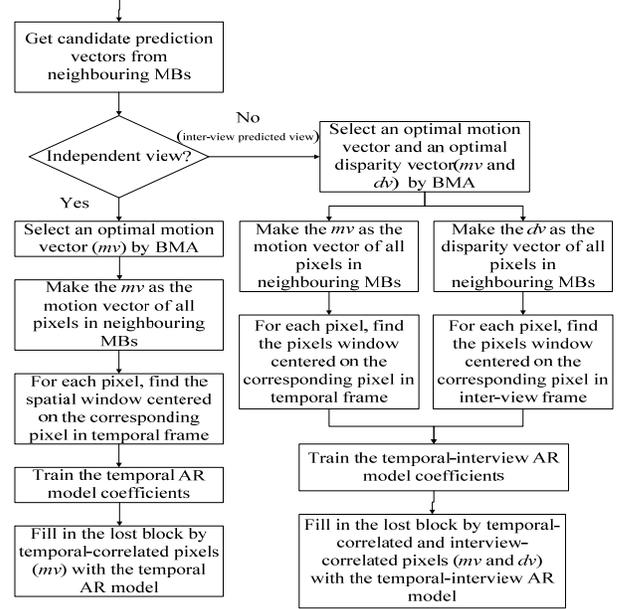

Fig. 2 Flowchart of the proposed auto-regressive based error concealment algorithm for stereoscopic video coding

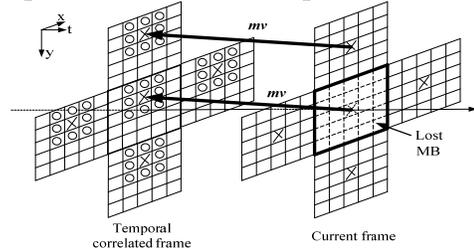### 2.2. Temporal AR model for independent view

Fig. 3 Temporal AR model for independent view

The proposed temporal AR model for independent view is shown in Fig.3. Suppose Xc be a region of m pixels in the current frame, and X_t be the region of pixels in the temporal-correlated frame along the temporal motion direction. For each pixel $X_c(i,j)$ in the region X_c located at (i, j), we find its temporal corresponding pixel in X_t, indicated by the temporal motion vector (mv_x, mv_y). Then we can approximate the current pixel as a weighted summation of pixels within a spatial window, centered on the

corresponding pixel in the temporal-correlated frame. The approximate value $\hat{X}_c(i,j)$ can be represented as

$$\hat{X}_c(i,j) = \sum_{u=-R}^{R} \sum_{v=-R}^{R} X_t\left(i+u+mv_x, j+v+mv_y\right) \cdot \alpha_{u,v} \qquad (2)$$

where $\alpha_{u,v}$ is the temporal AR coefficient located at (u, v), and R denotes the radius of the spatial window, thus the window size n is $(2R+1)\times(2R+1)$.

Due to the piecewise characteristics of nature image, we assume the AR coefficients $\alpha = (\alpha_{-R,-R}, \alpha_{-R,-R+1}, \ldots, \alpha_{R,R})^t$ remain the same for all the pixels in the region. Then, according to Eq. (2), we can obtain

$$\hat{X}_c = f(X_t)\alpha \qquad (3)$$

where $f(X_t)$ is a function which transfer X_t to a $m \times n$ dimensional matrix. For each pixel in X_t indicated by the temporal motion, the function gets a spatial window centered on the pixel.

Note that, for each pixel within the lost MB, we can approximate its value by the proposed AR model. The temporal motion has been selected by BMA, so we need to compute the AR coefficients for the lost MB.

Based on the piecewise characteristics of nature image, we assume the AR coefficients and the motion remain the same for the lost MB and the four neighboring MBs around the lost MB, as shown in Fig. 2. Thus utilizing the available neighboring MBs and the motion selected by BMA, we can compute the AR coefficients for the lost MB.

Mean squared error (MSE) criterion and the least square (LS) algorithm are utilized to compute the AR coefficient vector α. For all the available neighboring MBs we define the resulting MSE as

$$\varepsilon^2(\hat{X}_c) = E\left(\left\| X_c - \hat{X}_c \right\|^2\right) \qquad (4)$$

Here we select the AR coefficient vector α which results in the least $\varepsilon^2(\hat{X}_c)$ as the optimum coefficients. According to the least square (LS) algorithm, for each coefficient $\alpha_{u,v}$, we set

$$\frac{dE\left(\left\| X_c - \hat{X}_c \right\|^2\right)}{d\alpha_{u,v}} = \frac{dE\left(\left\| X_c - f(X_t)\alpha \right\|^2\right)}{d\alpha_{u,v}} = 0 \qquad (5)$$

Note that $f(X_t)$ is a $m \times n$ dimensional matrix, where m is the total number of pixels which is utilized for training AR model coefficients, and $n = (2R+1)\times(2R+1)$ is the size of spatial window. Then, according to Eq. (5) we can derive

$$\alpha = \left((f(X_t))^T f(X_t)\right)^{-1}(f(X_t))^T X_c \qquad (6)$$

Utilizing the coefficients derived by Eq. (6), we can easily recover the lost MB according to Eq. (3).

2.3. Temporal-interview AR model for inter-view predicted view

Fig. 4 depicts the proposed temporal-interview AR model for inter-view predicted view. Suppose X_v be the region of m pixels in the interview-correlated frame along the interview disparity direction. Then, applying the temporal-interview AR model, the approximate value $\hat{X}_c(i,j)$ can be represented as

$$\begin{aligned} \hat{X}_c(i,j) &= \sum_{u=-R}^{R} \sum_{v=-R}^{R} X_t\left(i+u+mv_x, j+v+mv_y\right) \cdot \alpha_{u,v} \\ &+ \sum_{u=-R}^{R} \sum_{v=-R}^{R} X_v\left(i+u+dv_x, j+v+dv_y\right) \cdot \beta_{u,v} \end{aligned} \qquad (7)$$

where $\alpha_{u,v}$ is the temporal AR coefficient located at (u, v), and $\beta_{u,v}$ is the interview AR coefficient located at (u, v).

Furthermore, we can obtain

$$\hat{X}_c = f(X_t)\alpha + g(X_v)\beta \qquad (8)$$

where $g(X_v)$ is a function which transfer X_v to a $m \times n$ dimensional matrix. For each pixel in X_v indicated by the interview disparity, the function gets a spatial window centered on the pixel.

Suppose $\Psi = \left(f(X_t), g(X_v)\right)$ be a $m \times 2n$ dimensional matrix, and $\omega = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (\alpha_{-R,-R}, \alpha_{-R,-R+1}, \ldots, \alpha_{R,R}, \beta_{-R,-R}, \beta_{-R,-R+1}, \ldots, \beta_{R,R})^t$ be the temporal-interview AR coefficient vector.

Then we obtain

$$\hat{X}_c = \Psi \omega \qquad (9)$$

Similarly as previous section, for each lost MB, utilizing the motion vector and disparity vector selected by BMA, and the available neighboring MBs, we can compute the temporal-interview AR coefficients for the lost MB. For all the available neighboring MBs, according to MSE criterion and the LS algorithm, we can derive

$$\omega = \left(\Psi^T \Psi\right)^{-1} \Psi^T X_c \qquad (10)$$

Utilizing the coefficients derived by Eq. (10), we can easily recover the lost MB according to Eq. (9).
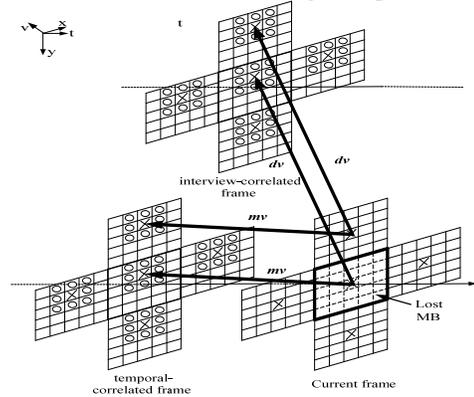

Fig. 4 Temporal-interview AR model for inter-view predicted view

3. SIMULATION RESULTS

We utilize the H.264/AVC reference software JM 10.0 to simulate a stereoscopic video coding system, as shown in Fig. 1. The left view is encoded independently with MCP, and the right view is predicted with both MCP and DCP. Then they are transmitted to the decoder respectively. The

video sequences *Race1* and *Rena* are encoded using the IPPP GOP structure for 240 frames within each view. Each row of MB composes a slice and is transmitted in a separate packet. The packet loss rates (PLR) at 5%, 10%, and 20% [9] are tested in experiments. The QP is set to 28 and the first I frame in a GOP in each view is error free for all tests. We compare our proposed stereoscopic AR error concealment scheme (AR) with temporal replacement (TR) method, the error concealment method of JM (JM), which only perform error concealment process in the temporal direction of a single view. For the proposed AR scheme, the radius of the spatial window R is set to be 1.

Firstly, we assume that the independent view sequence is transmitted correctly and the inter-view predicted view sequence is transmitted with packet loss. Under the condition, there's no error propagations from the independent view to the inter-view predicted view, thus we can test the performance of our proposed temporal-interview AR model for the inter-view predicted view. Table 1 shows the PSNR performance results of different error concealment schemes for the inter-view predicted view under the given conditions. AR has 3.70dB~6.70dB error concealment performance improvement than TR, 1.58dB~3.32dB improvement than JM. The results indicate that the proposed AR scheme can utilize inter-view correlations to improve the error concealment quality for the inter-view predicted view.

For both the independent and inter-view predicted view sequences are transmitted with the same PLR, Table 2 shows the PSNR performance comparison of the three different methods. AR has 3.71dB~6.05dB error concealment performance improvement than TR, 0.62dB~1.23dB improvement than JM.

Table 1 Average PSNR (dB) performance comparison of different error concealment schemes for the inter-view predicted view

Video sequence	PSNR (dB)	Packet loss rate		
		5%	10%	20%
Rena	TR	35.30	32.79	30.24
	JM	36.85	34.57	32.26
	AR	**39.00**	**37.32**	**35.58**
Race1	TR	29.58	27.07	25.00
	JM	33.82	31.75	29.53
	AR	**35.40**	**33.55**	**31.70**

4. CONCLUSIONS

Robust stereoscopic video transmission has posed a technical challenge for stereoscopic video coding. In this paper, combining the temporal and interview correlation in stereoscopic video and the superior property of the AR model, we propose an auto-regressive model based error concealment scheme which is composed of a temporal AR model for independent view and a temporal-interview AR model for inter-view predicted view. Applying the proposed

AR model, each pixel of the lost block is interpolated as a weighted summation of pixels in the reference frames along the prediction directions. Motions and disparities for lost blocks are selected by the boundary matching algorithm. And the proposed AR model coefficients are computed according to the spatial neighboring pixels and their temporal-correlated and interview-correlated pixels indicated by the selected best predicted directions. Simulation results show that the proposed algorithm can improve the performance of reconstructed stereoscopic video sequences.

Table 2 Average PSNR (dB) performance comparison of different error concealment schemes

Video sequence	PSNR (dB)	Packet loss rate		
		5%	10%	20%
Rena	TR	32.84	29.57	27.43
	JM	35.84	32.99	30.63
	AR	**36.55**	**33.61**	**31.37**
Race1	TR	28.49	25.71	23.98
	JM	33.17	30.53	28.52
	AR	**34.26**	**31.76**	**29.64**

5. REFERENCES

[1] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets", *Proc. IEEE ICASSP 86*, pp. 521 – 524, 1986.

[2] Aign S. and Fazel K., "Temporal and spatial error concealment technique for hierarchical MPEG-2 video codec", *IEEE International Conference on Communications*, Vol. 3, pp.1778-1783, June 1995.

[3] W. M. Lam, A. R. Reilbman and B. Liu, "Recovery of lost or erroneously received motion vectors", *Proc. IEEE ICASSP 93*, pp. V417-V420, April 1993.

[4] S. Tsekeridou and I. Pitas, "MPEG-2 error concealment based on block-matching principles", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 646-658, June 2000.

[5] Mei-Juan Chen, Liang-Gee Chen and Ro-Min Weng, "Error concealment of lost motion vectors with overlapped motion compensation"**,** *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, Issue 3, pp. 560-563, June 1997.

[6] T. Chen, X. Zhang, and Y. Q. Shi, "Error concealment using refined boundary matching algorithm", *International Conference on Information Technology: Research and Education*, pp.55-59 Aug. 2003.

[7] S. Knorr, C. Clemens, M. Kunter and T. Sikora, "Robust concealment for erroneous block bursts in stereoscopic images", *Proc. of 3DPTV2004*, pp.820-827, Sept. 2004.

[8] X. Wu, K. U. Barthel, and W. Zhang, "Piecewise 2D autoregression for predictive image coding", *Proc. IEEE International Conference on Image Processing,* pp. 901-904, 1998.

[9] S. Wenger, "Error patterns for internet experiments", ITU-T SG16 Doc. Q15-I-16r1, 1999.