Auto-Regressive Model Based Error Concealment Scheme for Stereoscopic Video Coding

Presenter: Xinguang Xiang
Authors: Xinguang Xiang, Debin Zhao,
Siwei Ma and Wen Gao
Harbin Institute of Technology&Peking University

Outline

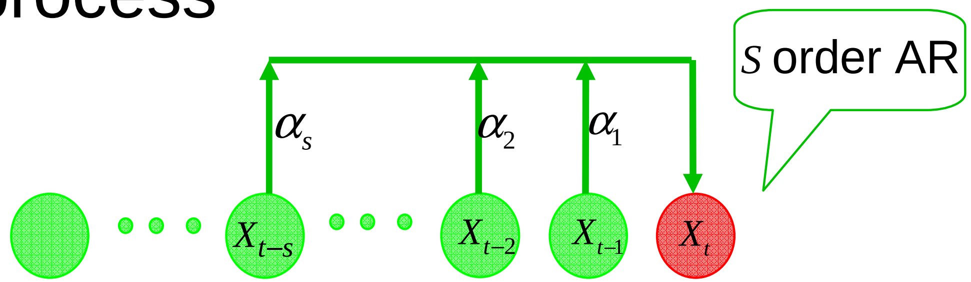
- Introduction
- Proposed auto-regressive model based error concealment scheme for stereoscopic video coding
- Experimental results and performance comparison
- Conclusions

Introduction

- Stereoscopic video coding
  - Provide viewers the 3-D sensation
  - Key: inter-view correlation
  - Prediction structure: disparity compensation prediction (DCP) in the inter-view direction and motion compensation prediction (MCP) in the temporal direction
- Error concealment
  - Purpose: robust video transmission over error-prone networks, reconstruct lost video contents, decoder side
  - Key: temporal correlation for single view coding
  - Aims at this problem for stereoscopic video coding

Auto-regressive model

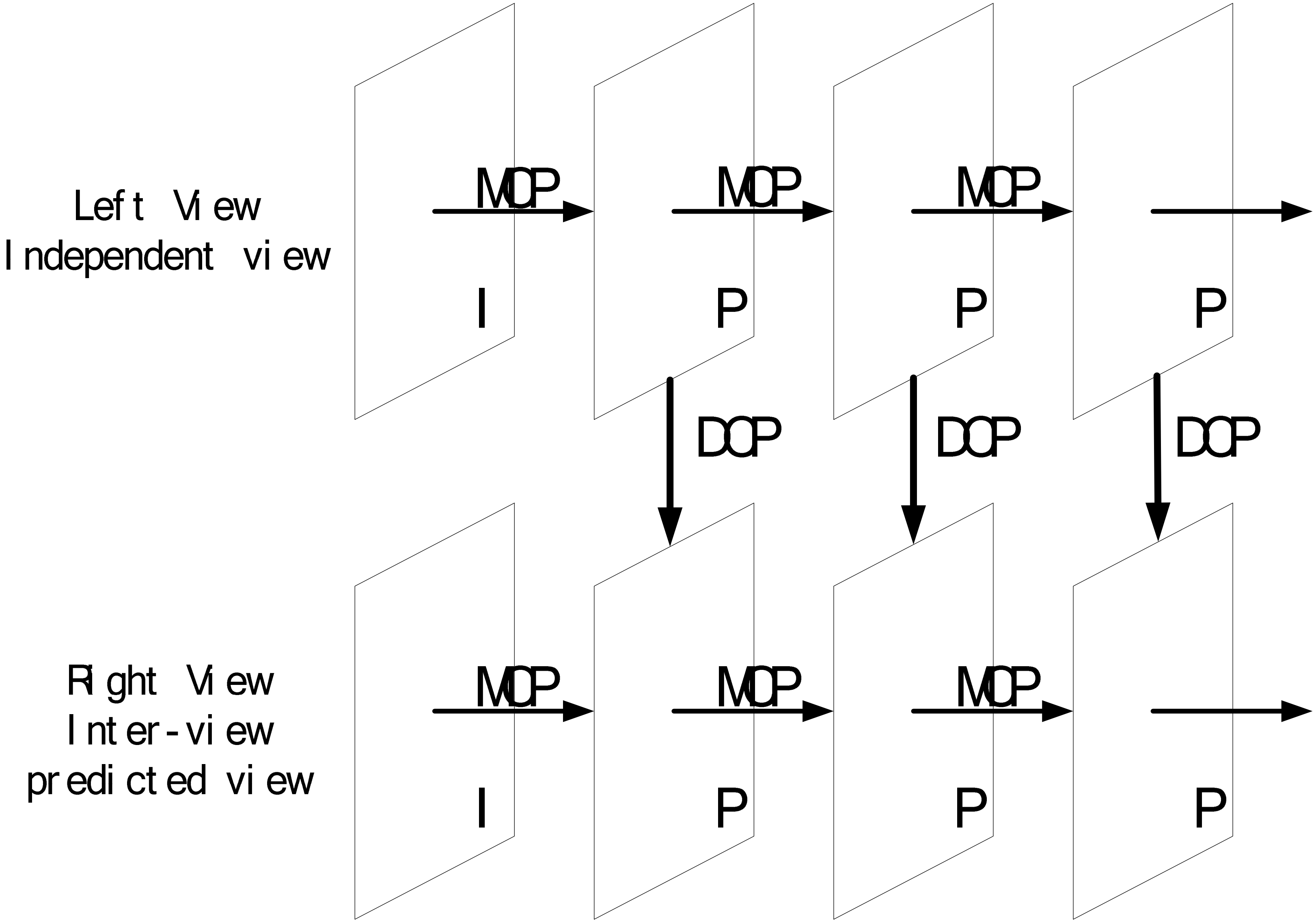
- Auto regressive (AR) model is an efficient description of random process



$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_s X_{t-s}$$

- AR has great ability of predicting future data
- It will be desirable to employ AR in error concealment for stereoscopic video coding

Basic prediction structure of two-view based stereoscopic video coding



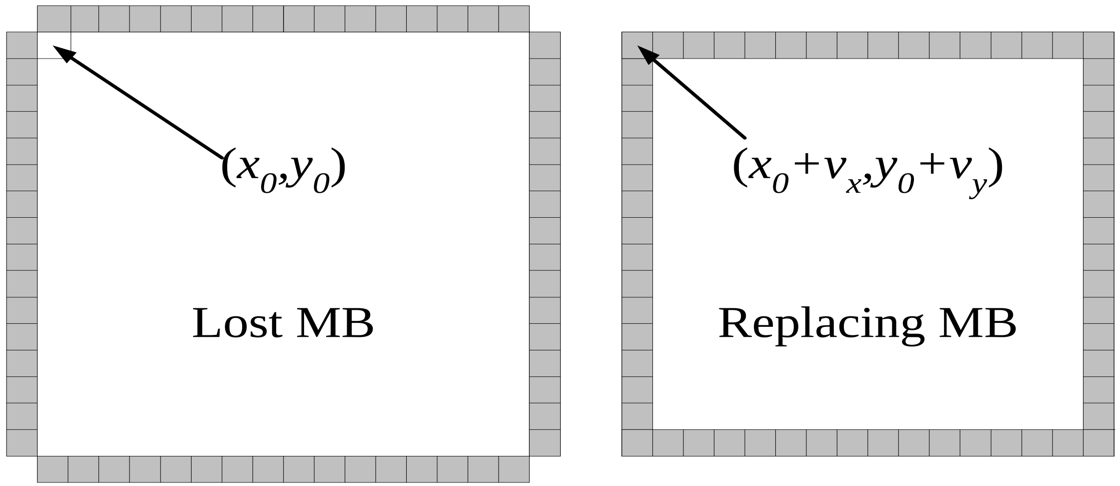
Proposed auto-regressive model based error concealment scheme for stereoscopic video coding

- Temporal AR model for independent view
  - Utilizing temporal motion information to establish AR model
  - Utilizing temporal AR model to reconstruct the lost video contents
- Temporal-interview AR model for inter-view predicted view
  - Utilizing temporal motion information and inter-view disparity information to establish AR model
  - Utilizing temporal-interview AR model to reconstruct the lost video contents

Selection of prediction directions

- In order to choose appropriate prediction vectors which are used as the prediction directions for AR model, we utilize the well known boundary matching algorithm (BMA) criterion.

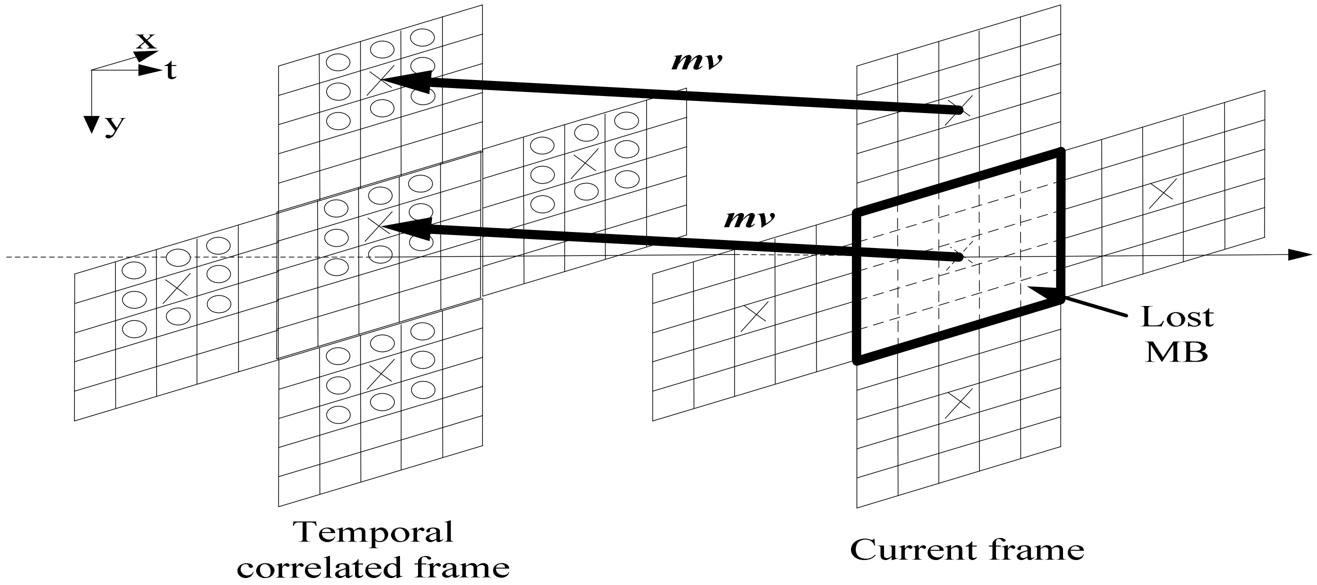
The cost function of BMA is defined as the absolute difference between the external boundary of the lost MB in the current frame and the internal boundary of the replacing MB in the reference frame



Current frame Reference frame

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

Temporal AR model for independent view



- 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

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

Temporal AR model for independent view

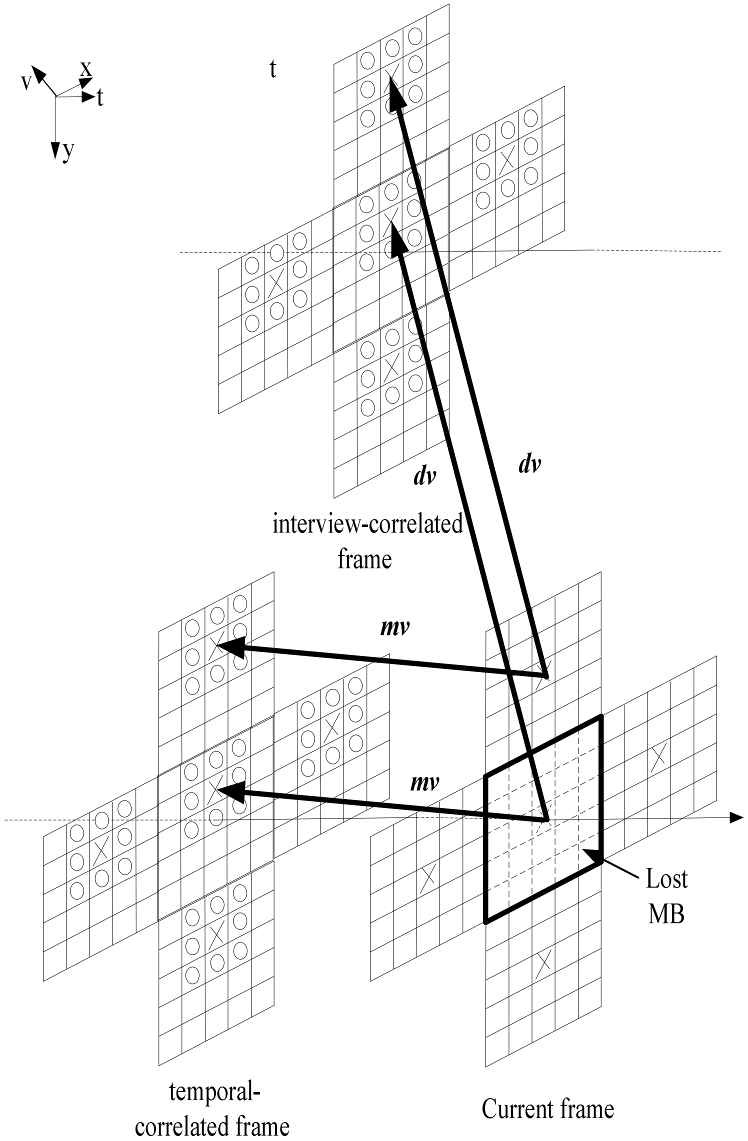
- Based on the piecewise characteristics of nature image, we assume the AR coefficients and the motion remain the same for the lost MB and the four neighboring MBs around the lost MB
- Mean squared error (MSE) criterion and the least square (LS) algorithm are utilized to compute the AR coefficient vector.

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

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

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

Temporal-interview AR model for inter-view predicted view



- We can approximate the current pixel as a weighted summation of pixels within two spatial windows, centered on the corresponding pixel in the temporal-correlated frame and interview-correlated frame

$$\hat{X}_c(i,j) = \sum_{u=-R}^{R} \sum_{v=-R}^{R} X_t(i+u+mv_x, j+v+mv_y) \cdot \alpha_{u,v} + \sum_{u=-R}^{R} \sum_{v=-R}^{R} X_v(i+u+dv_x, j+v+dv_y) \cdot \beta_{u,v}$$

Experimental Results

- Simulation conditions
  - Tools: JM 10.0
  - Left view: MCP; right view: MCP and DCP
  - Anchor: simulated stereoscopic video coding system
  - Temporal replacement (TR) algorithm, error concealment tools in JM (JM) and our proposed algorithm (AR) are simulated
  - Sequences: *Ballroom* and *Race1*
  - GOP structure: IPPP for 240 frames
  - The packet loss rates (PLR): 5%, 10%, 20%

PSNR performance comparison

- Average PSNR (dB) comparison for the right view sequence supposing the left view is error free
- AR has 3.70dB~6.70dB error concealment performance improvement than TR, 1.58dB~3.32dB improvement than JM.

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**

PSNR performance comparison

- Both the left and right view sequences are transmitted with packet loss
- AR has 3.71dB~6.05dB error concealment performance improvement than TR, 0.62dB~1.23dB improvement than JM

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**

Conclusions

- An auto-regressive model based error concealment scheme is proposed for stereoscopic video coding
  - Combining the temporal and interview correlation in stereoscopic video and the superior property of the AR model
  - A temporal AR model for independent view and a temporal-interview AR model for inter-view predicted view
  - Utilizing AR model to reconstruct the lost video contents

- Q&A

Thank you!