Schemes for multiple description coding of stereoscopic video

Andrey Norkin[1], Anil Aksay[2], Cagdas Bilen[2], Gozde Bozdagi Akar[2], Atanas Gotchev[1], and Jaakko Astola[1]

[1] Tampere University of Technology
Institute of Signal Processing
P.O.Box 553, FIN-33101 Tampere, FINLAND,
{andrey.norkin, atanas.gotchev, jaakko.astola}@tut.fi
[2] Middle East Technical University, Ankara, Turkey
{anil, cbilen, bozdagi}@eee.metu.edu.tr

Abstract. This paper presents and compares two multiple description schemes for coding of stereoscopic video, which are based on H.264. The SS-MDC scheme exploits spatial scaling of one view. In case of one channel failure, SS-MDC can reconstruct the stereoscopic video with one view low-pass filtered. SS-MDC can achieve low redundancy (less than 10%) for video sequences with lower inter-view correlation. MS-MDC method is based on multi-state coding and is beneficial for video sequences with higher inter-view correlation. The encoder can switch between these two methods depending on the characteristics of video.

1 Introduction

Recently, as the interest in stereoscopic and multi-view video has grown, different video coding methods are investigated. *Simulcast* coding is coding the video from each view as monoscopic video. *Joint* coding is coding the video from all the views jointly to exploit correlation between different views. For example, left sequence is coded independently, and frames of the right sequence are predicted from either right or left frames. A multi-view video coder (MMRG) has been proposed in [1]. This coder has several operational modes corresponding to different prediction schemes for coding of stereoscopic and multi-view sequences. MMRG coder is based on H.264, which is the current state-of-the-art video coder. This coder exploits correlation between different cameras in order to achieve higher compression ratio than the simulcast coding.

Compressed video sequence is vulnerable to transmission errors. This is also true for stereoscopic video. Moreover, due to more complicated structure of the prediction path errors in the left sequence can propagate further in the subsequent left frames and also in the right frames.

One of the popular methods providing error resilience to compressed video is multiple description coding (MDC) [2]. MDC has a number of similarities to coding of stereoscopic video. In MDC, several bitstreams (descriptions) are generated from the source information. The resulting descriptions are correlated and have similar importance. Descriptions are independently decodable at basic quality level. The more descriptions are received, the better is reconstruction quality. MDC is especially beneficial when combined with multi-path transport [3], i.e. when each description is sent to the decoder over a different path.

In simulcast coding, bitstream from each view can be independently decoded with target quality to obtain monoscopic video. When both views are decoded, stereoscopic video is obtained. Simulcast coding has higher bitrate than joint coding. However, simulcast coding cannot provide stereoscopic reconstruction if one sequence is lost. Thus, one can think of exploiting the nature of stereoscopic video in order to design a reliable MD stereoscopic video coder. However, to our knowledge, there has not been any extensive research on MDC for stereo- and multi-view video coding.

In this paper, we present two MDC approaches for stereoscopic video. These approaches produce balanced descriptions and are able to provide stereoscopic reconstruction in case of one channel failure for the price of moderate coding redundancy. The approaches are referred to as Scaling Stereo-MDC (SS-MDC) and Multi-state Stereo-MDC (MS-MDC). Both the proposed methods are drift-free and can be used interchangeably.

2 Spatial scaling stereo-MDC scheme

There are two theories about the effects of unequal bit allocation between left and right video sequences. Those theories are *fusion theory* and *suppression theory* [4], [5], [6]. In fusion theory, it is believed that total bit budget should be equally distributed between two views. According to suppression theory, the overall perception in a stereo-pair is determined by the highest quality image. Therefore, one can compress the target image as much as possible to save bits for the reference image, so that overall distortion is the lowest.

Our SS-MDC approach is based on these two theories. In [7], the perception performance of spatial and temporal down-scaling for stereoscopic video compression has been studied. The obtained results indicate that spatial and spatiotemporal scaling provide acceptable perception performance with a reduced bitrate. It gave us the idea of using scaled stereoscopic video as side reconstruction in our MD coder.

2.1 Prediction scheme

Fig. 1 presents the scheme exploiting spatial scaling of one view (SS-MDC). In *Description 1*, left frames are predicted only from left frames, and right frames are predicted from both left and right frames. Left frames are coded with the original resolution; right frames are downsampled prior to encoding. In *Description 2*, right frames are coded with the original resolution and left frames are downsampled.

When both descriptions are received, left and right sequences are reconstructed in full resolution. If one description is lost due to channel failures, the

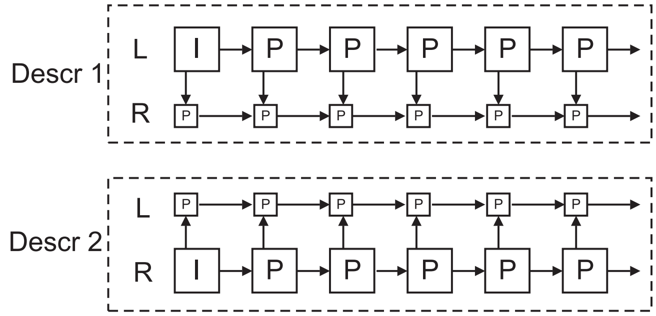


Fig. 1. MDC scheme based on spatial scaling (SS-MDC).

decoder reconstructs a stereoscopic video pair, where one view is low-pass filtered. A stereo-pair where one view has the original resolution and another view is low-pass filtered provides acceptable stereoscopic perception. After the channel starts working again, the decoding process can switch back to the central reconstruction (where both views have high resolution) after the IDR picture is received.

The proposed scheme can easily be done standard compatible. If each description is coded with standard compatible mode of MMRG coder [1] then standard H.264 decoder can decode the original resolution sequence from each description. The proposed scheme produces balanced descriptions as left and right sequences usually have similar characteristics and are encoded with the same bitrate and visual quality. The proposed SS-MDC scheme is drift-free, i.e. it does not introduce any mismatch between the states of the encoder and decoder in case of description loss.

2.2 Downsampling

Downsampling consists of low-passed filtering followed by decimation. The following filters are used:

13-tap downsampling filter:
$\{0, 2, 0, -4, -3, 5, 19, 26, 19, 5, -3, -4, 0, 2, 0\}/64$
11-tap upsampling filter:
$\{1, 0, -5, 0, 20, 32, 20, 0, -5, 0, 1\}/64$

Filters are applied to all Y,U, and V channels in both horizontal and vertical directions, and picture boundaries are padded by repeating the edge samples. These filters are used in Scalable Video Coding extention of H.264 [8] and explained in [9]. The downscaling is done by factors of 2 in both dimensions. In motion estimation of the downscaled sequence, frames with the original resolution are also scaled by the same factor for proper estimation.

2.3 Redundancy of SS-MDC

The bitrate generated by the SS-MDC coder is $R = R^* + \rho_{sim} + \rho_d$, where R^* is the bitrate obtained with the single description coding scheme providing the best compression, ρ_{sim} is the redundancy caused by using simulcast coding instead of joint coding, and ρ_d is the bitrate spent on coding of the downscaled sequences. Thus, the redundancy $\rho = \rho_{sim} + \rho_d$ of the proposed method is bounded by the redundancy of the simulcast coding ρ_{sim}. The redundancy of the simulcast coding ρ_{sim} depends on characteristics of the video sequence and varies from one sequence to another. The redundancy ρ_d of coding two downsampled sequences can be adjusted to control the total redundancy ρ. Redundancy ρ_d is adjusted by changing scaling factor (factors of two in our implementation) and quantization parameter QP of the downscaled sequence.

3 Multi-state stereo-MDC scheme

The MS-MDC scheme is shown in Fig. 2. Stereoscopic video sequence is split into two descriptions. Odd frames of both left and right sequences belong to *Description 1*, and even frames of both sequences belong to *Description 2*. Motion compensated prediction is performed separately in each description. In *Description 1*, left frames are predicted from preceding left frames of *Description 1*, and right frames are predicted from preceding right frames of *Description 1* or from the left frames corresponding to the same time moment. The idea of this scheme is similar to video redundancy coding (VRC) [10] and multi-state coding [11].

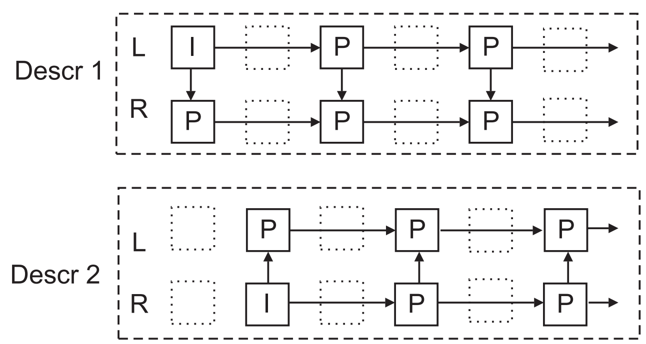


Fig. 2. Multistate stereo MDC.

If the decoder receives both descriptions, the original sequence is reconstructed with the same frame rate. If one description is lost, stereoscopic video is reconstructed with the half of the original frame rate. Another possibility is to employ a frame concealment technique for the lost frames. As one can see from Fig. 2, missed (e.g. odd) frame can be concealed by employing motion vectors of the next (even) frame, which uses only previous even frame as a reference for motion-compensated prediction .

This MDC scheme does not allow to adjust coding redundancy. However, for some video sequences it allows to reach bitrates lower than bitrate of the simulcast coding $R_{sim} = R^* + \rho_s$. This method can be easily generalized for more than two descriptions. MS-MDC also does not introduce any mismatch between the states of the encoder and decoder in case of description loss.

4 Simulation results

In the experiments, we compare side reconstruction performance of the proposed MDC schemes. The results are provided for four stereoscopic video pairs: Traintunnel (720×576, 25 fps, moderate motion, separate cameras), Funfair (360×288, 25 fps, high motion, separate cameras), Botanical (960×540, 25 fps, low motion, close cameras) and Xmas (640×480, 15 fps, low motion, close cameras). Both algorithms are applied to these videos. In all the experiments, I-frames are inserted every 25 frames.

The reconstruction quality measure is PSNR. PSNR value of a stereo-pair is calculated according to the following formula, where D_l and D_r represent the distortions in the left and right frames [12].

$$PSNR_{pair} = 10 \log_{10} \frac{255^2}{(D_l + D_r)/2}$$

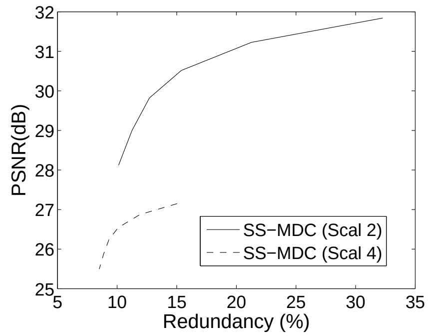
In the experiments, average $PSNR_{pair}$ is calculated over the sequence. Redundancy is calculated as the percentage of additional bitrate over the encoding with the minimal bitrate R^*, i.e. the bitrate of a *joint* coding scheme.

To show characteristics of the video sequences, we code them by *joint* coder and *simulcast* coder for the same PSNR. The results are shown in Table 1. The experiments for MD coding use the same values of the D_0 and R^*, which are given in the Table 1. One can see that Traintunnel and Funfair sequences show low inter-view correlation, and sequences Botanical and Xmas show high inter-view correlation. Thus, Botanical and Xmas have high redundancy of simulcast coding ρ_{sim}, which is the lower bound for redundancy of SS-MDC coding scheme.

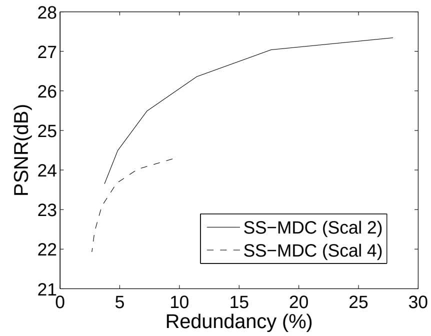
The SS-MDC scheme is tested for downsampling factors of 2 and 4 in both vertical and horizontal directions. For each downscaling factor, we change quantization parameter (QP) of the downscaled sequence to achieve different levels of redundancy.

Sequence	D_0, dB	$R^* = R_{joint}$, Kbps	R_{sim}, Kbps	ρ_{sim}, %
TrainTunnel	35.9	3624	3904	7.7
Funfair	34.6	3597	3674	2.2
Botanical	35.6	5444	7660	40.7
Xmas	38.7	1534	2202	43.5

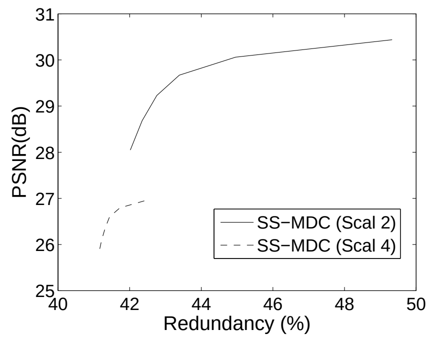
Table 1. Joint and simulcast coding.



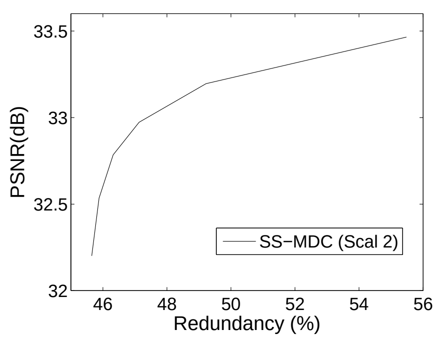
(a) Traintunnel. MS-MDC: $D_1 = 30.7$ dB, $\rho = 41.4\%$.



(b) Funfair. MS-MDC: $D_1 = 26.8$ dB, $\rho = 24.3\%$.



(c) Botanical. MS-MDC: $D_1 = 31.4$ dB, $\rho = 28.3\%$.



(d) Xmas. MS-MDC: $D_1 = 29.6$ dB, $\rho = 30.1\%$.

Fig. 3. Redundancy rate-distortion curves for test sequences.

The results for the second scheme (MS-MDC) are given only for one level of redundancy. The reason is that this method does not allow to adjust redundancy since the coding structure is fixed as in Figure 2. The redundancy of MS-MDC method takes only one value and is determined by characteristics of the video sequence.

Fig. 3 shows the redundancy-rate distortion (RRD) curves [13] for SS-MDC and the values for MS-MDC for test sequences. The results are presented as PSNR of a side reconstruction (D_1) vs redundancy ρ. The results for SS-MDC are given for scaling factors 2 and 4. For sequence Xmas, simulation results for scaling factor 4 are not shown, as PSNR is much lower than for scaling factor 2.

The simulation results show that reconstruction from one description can provide acceptable video quality. The SS-MDC method can perform in a wide range of redundancies. Downscaling with factor 2 provides good visual quality with acceptable redundancy. However, the performance of SS-MDC depends to

Sequence	Joint	SS-MDC	MS-MDC
Traintunnel	0.94	0.78	0.90
Funfair	0.92	0.80	0.85
Botanical	0.65	0.60	0.63
Xmas	0.66	0.56	0.61

Table 2. Fraction of MVs in the right sequence which point to previous right frames.

a great extent on the nature of stereoscopic sequence. This method can achieve very low redundancy (less than 10%) for sequences with lower inter-view correlation (Traintunnel, Funfair). However, it has higher redundancy in stereoscopic video sequences with higher inter-view correlation (Xmas, Botanical). The perception performance of SS-MDC is quite good as the stereo-pair perception is mostly determined by quality of the high-resolution picture.

The MS-MDC coder perform usually with 30-50% redundancy and can provide acceptable side reconstruction even without error concealment algorithm (just by copying the previous frame instead of the lost frame). MS-MDC should be used for sequences with higher inter-view correlation, where SS-MDC shows high redundancy.

The encoder can decide which scheme to use by collecting the encoding statistics. Table 2 shows the statistics of motion vectors (MVs) prediction for *joint* coding mode, SS-MDC, and MS-MDC. The statistics are collected for P-frames of the right sequence. Values in Table 2 show the fraction of motion vectors m which point to the frames of the same sequence, i.e. the ratio of motion vectors to sum of the motion and disparity vectors in the right sequence frames. One can see that the value m correlates with the redundancy of *simulcast* coding ρ_{sim} given in Table 1. The value m could tell the decoder when to switch from SS-MDC to MS-MDC and vice versa.

Thus, the encoder operates as follows. Once the encoding mode has been chosen depending on m, the encoding process starts, and the statistics are being collected. Before the encoding IDR picture, encoder compares the value m of the recent N frames with threshold 0.7 and decides whether to switch to a different mode or not. Thus, the encoder adaptively chooses SS-MDC or MS-MDC mode depending on characteristics of the video sequence.

5 Conclusions and future work

Two MDC approaches for stereoscopic video have been introduced. These approaches produce balanced descriptions and provide stereoscopic reconstruction with acceptable quality in case of one channel failure for the price of moderate redundancy (in the range of 10-50%). Both the presented approaches provide drift-free reconstruction in case of description loss.

The performance of these approaches depends on characteristics of stereoscopic video sequence. The approach called SS-MDC performs better for sequences with lower inter-view correlation while MS-MDC approach performs better for sequences with higher inter-view correlation. The criterium for switching between the approaches is used by the encoder to choose the approach that provides better performance for this sequence.

Our plans for future research are optimization of the proposed approaches and study of their performance over transmission channel, such as DVB-H transport.

6 Acknowledgements

This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

References

1. Bilen, C., Aksay, A., Bozdagi Akar, G.: A multi-view video codec based on H.264. In: Proc. IEEE Conf. Image Proc. (ICIP), Oct. 8-11, Atlanta, USA (2006)
2. Wang, Y., Reibman, A., Lin, S.: Multiple description coding for video delivery. Proceedings of the IEEE **93** (2005) 57–70
3. Apostolopoulos, J., Tan, W., Wee, S., Wornell, G.: Modelling path diversity for multiple description video communication. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing. Volume 3. (2002) 2161–2164
4. Julesz, B.: Foundations of cyclopeon perception. The University of Chicago Press (1971)
5. Dinstein, I., Kim, M.G., Henik, A., Tzelgov, J.: Compression of stereo images using subsampling transform coding. Optical Engineering **30** (1991) 1359–1364
6. Woo, W., Ortega, A.: Optimal blockwise dependent quantization for stereo image coding. IEEE Trans. on Cirquits Syst. Video Technol. **9** (1999) 861–867
7. Aksay, A., Bilen, C., Kurutepe, E., Ozcelebi, T., Bozdagi Akar, G., Civanlar, R., Tekalp, M.: Temporal and spatial scaling for stereoscopic video compression. In: Proc. EUSIPCO'06, Sept. 4-8, Florence, Italy (2006)
8. Reichel, J., Schwarz, H., Wien, M.: Scalable video coding - working draft 3. In: JVT-P201, Poznan, PL, 24-29 July. (2005)
9. Segall, S.A.: Study upsampling/downsampling for spatial scalability. In: JVT-Q083, Nice, FR, PL, 14-21 October. (2005)
10. Wenger, S., Knorr, G., Ott, J., Kossentini, F.: Error resilience support in H.263+. IEEE Trans. Circuits Syst. Video Technol. **8** (1998) 867–877
11. Apostolopoulos, J.: Error-resilient video compression through the use of multiple states. In: Proc. Int. Conf. Image Processing. Volume 3. (2000) 352–355
12. Boulgouris, N.V., Strintzis, M.G.: A family of wavelet-based stereo image coders. IEEE Trans. on Cirquits Syst. Video Technol. **12** (2002) 898–903
13. Orchard, M., Y.Wang, Vaishampayan, V., Reibman, A.: Redundancy rate distortion analysis of multiple description image coding using pairwise correlating transforms. In: Proc. Int. Conf. Image Processing, Santa Barbara, CA (1997) 608–611