SHARING OF MOTION VECTORS IN 3D VIDEO CODING

Stefan Grewatsch and Erika Müller

University of Rostock
Institute of Communications Engineering
Richard-Wagner-Str. 31
18119 Rostock, Germany

ABSTRACT

Current investigations in the area of 3D video systems are based on depth map sequences. They are used to determine the compatibility with conventional 2D video systems. Instead of transmitting or storing two 2D video sequences for each eye, 2D video and the appropriate depth information is delivered. If the viewer owns a standard TV set, only the video is displayed. In the case of additionally installed hardware, e.g. stereoscopic display or shutter glasses, a second virtual view is computed based on the depth information. The amount of data for transmitting or storing the additional depth information is limited to 20 percent of the MPEG-2 coded 2D video data stream. This paper shows an approach for compression of the depth map sequences based on MPEG-2 within the predetermined limit by sharing the vectors for motion compensation with the 2D video data.

1. INTRODUCTION

The project *Advanced Three-Dimensional Television System Technologies* (ATTEST) shows an evolutionary introduction of depth perception into the existing 2D digital television framework. It is a part of the European Information Society Technologies (IST) programme and has the aim of designing a backwards-compatible and flexible 3D television system [1]. Instead of transmitting or storing two video sequences for each eye, monoscopic video and associated per-pixel depth information is delivered.

A key feature of the outlined 3D video chain will be the flexibility for the consumer's display. If the viewer owns a standard TV set the 2D video is displayed only. In case of additionally installed hardware, e.g. stereoscopic display or shutter glasses, a second virtual view is synthesized based on the depth information. This can be done by using image-based rendering (IBR) methods, which use the per-pixel depth information to warp the original image points into the desired view [2].

The depth information is stored as second video sequence called a *depth map sequence* containing only one color component. The size of the depth maps is the same as the video frames. By using the depth values for every color pixel of the video frame, the position in 3D space is determined. Due to the limited transmission bandwidth and storage capacity, the depth map sequence has to be compressed, too. Therefore common video compression methods for reducing spatial and temporal redundancy can be used. In order to reduce temporal redundancy, motion estimation and compensation is applied. This fundamental method transmits a displacement vector and an associated residual data block. Due to the relationship between the video and the depth map sequence, in this paper the approach of using the same motion vectors for compression of both sequences is analyzed. Especially at lower bitrates the coding cost for motion vectors outweighs that one for coding the residual data blocks. In order to save transmission bandwidth or storage capacity the motion vectors are shared during compression and transmitted only once.

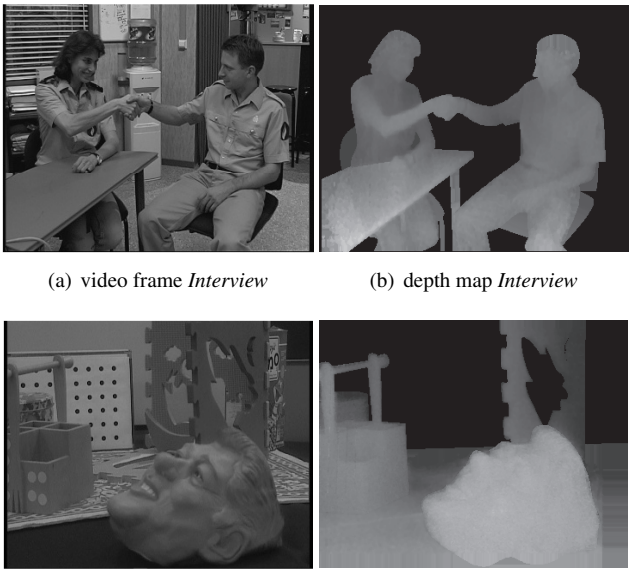
The video compression standard MPEG-2 [3] has been established for application scenarios like digital television (DVB) and storage media (DVD). Therefore this paper is focused on MPEG-2 for coding the video and depth map sequences. The approach of sharing the motion vectors is based on that standard, too.

The paper is organized as follows. First the motion vector fields of the 2D videos and the associated depth map sequences are analyzed. Assuming that the horizontal and vertical vector components of both fields have approximately the same values the correlation is investigated. Afterwards the proposed method of motion vector sharing is explained. In order to evaluate the approach coding experiments are carried out based on the MPEG-2 test model 5 (TM5) implemented by the *MPEG Software Simulation Group*. Finally, the experimental results are presented and discussed.

2. ANALYSIS OF MOTION VECTOR FIELDS

The video and the depth map sequence describe the content of the same scene. The video frames contain the texture details like color and surface structure and the depth maps

0-7803-8554-3/04/$20.00



(a) video frame *Interview* (b) depth map *Interview*

(c) video frame *Orbi* (d) depth map *Orbi*

Fig. 1. 3D video consisting of conventional video and depth maps, test sequences *Interview* and *Orbi* [4]

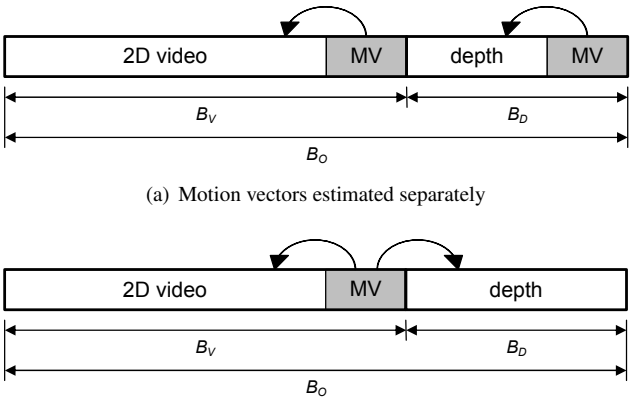
store the information about the distance of the scene objects. If we look closely at Fig. 1 we recognize the objects silhouette in both. Within consecutive video frames and depth maps the scene objects are moving in the same direction. This leads to the assumption that the motion vectors for temporal prediction could be shared in both sequences. Consequently, they have to be coded only once. If this approach is applied either the overall bitrate B_O of the 3D video sequence could be reduced or the coding cost for one motion vector set could be used for frame or residual coding, too. In Fig. 2(b) the latter case is outlined, Fig. 2(a) illustrates independent compression of both sequences.

In order to verify the above-described assumption, the correlation between the motion vector fields of video and depth map sequences were analyzed. Using the MPEG-2 reference software [5] the estimated motion vector fields from both sequences based on uncoded reference frames were extracted. In addition the parameter configuration C_1 was used (see section 3). By applying eqn. (1) the two-dimensional correlation coefficient r for the vertical and horizontal components were calculated.

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \bar{A}\right)\left(B_{mn} - \bar{B}\right)}{\sqrt{\left[\sum_m \sum_n \left(A_{mn} - \bar{A}\right)^2\right]\left[\sum_m \sum_n \left(B_{mn} - \bar{B}\right)^2\right]}}, \quad (1)$$

$$\bar{A} = \frac{1}{mn}\sum_m \sum_n A_{mn}, \quad \bar{B} = \frac{1}{mn}\sum_m \sum_n B_{mn}$$

The range of values for r is $-1 \leq r \leq 1$. The matrices A and



(a) Motion vectors estimated separately

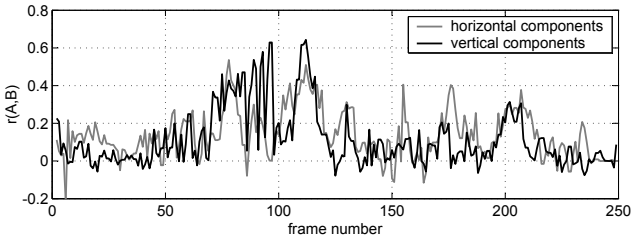
(b) Sharing of 2D video motion vectors

Fig. 2. Different kinds of using motion vectors

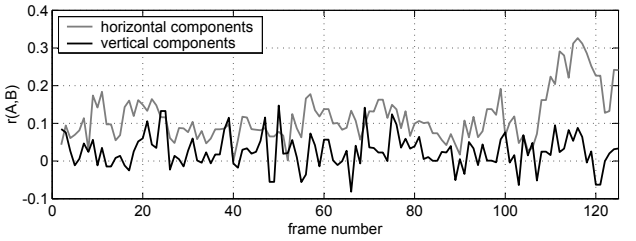
B contain the vertical or horizontal components of the motion vector fields of one 2D video frame and one depth map. If the correlation between the field components is strong the coefficient r will be nearly one. Otherwise if the correlation is small r will be close to zero.

In Fig. 3(a) and Fig. 3(b) the calculation results for the 3D video test sequences *Interview* and *Orbi* are shown. As expected a correlation between the motion vector fields does exist. When the motion of scene objects is faster, the correlation is stronger between the vector fields. The sequence *Interview* shows two persons that are shaking hands from frame 75 to 120. While giving and releasing hands, the horizontal and vertical components of the motion vectors are significantly correlated. While shaking hands the vertical components are more closely correlated. In the sequence *Orbi* the camera moves horizontally around the scene setup. The motion is slow during most of the frames. At the end it increases and results in a closer correlation between the horizontal components of both motion vector fields. The vertial components are not significantly correlated.

Encouraged by the preliminary investigations coding experiments were executed. For reasons of integrity it has to be pointed out that the aim of motion estimation is to reduce the prediction residual for a specific macro block and thus the coding cost. There is no intention of estimating the real motion in the scene. The motion vectors are specific to the frame content. Sharing them may be suboptimal in respect to the prediction error criterion. Because of the higher priority of the video sequence its estimated motion vectors were shared. Therefore it has to be expected that there is diminished compression efficiency at higher bitrates for the depth map sequence. At lower bitrates the transform coefficients are quantized strongly. Consequently, the coding cost for the motion vectors outweighs that of the prediction residuals. Because the motion vectors were coded within the 2D video bit stream, the whole bitrate B_D can be used



(a) Test sequence *Interview*



(b) Test sequence *Orbi*

Fig. 3. Correlation coefficient between the motion vector fields of video and depth map sequences

for the depth maps frame and residual coding. Therefore it is possible to transmit any depth map information especially at lower bitrates.

3. INVESTIGATION OF CODING GAIN

In the following section the investigation of the effective coding gain by using motion vector sharing is described. The test sequences which were used are interlaced. For efficient coding MPEG-2 provides the option of handling the top and bottom field as consecutive frames with half the vertical resolution [3]. If there are fast moving objects in the scene or the camera is panning, the whole frame contains high spatial frequencies in vertical direction. Thus motion compensation and transform coding will not be efficient. Another option is to switch adaptively at macro block level between frame and field prediction. The latter is further supported by an adaptive DCT coding. Due to subsampling the fields have higher spatial variance in vertical direction. This results in an asymmetrical distribution of the transform coefficients in frequency domain. Therefore an alternative sorting for run length coding is used. In Table 1 the parameter configurations used for coding are shown. As is common one group of pictures (GOP) consists of 12 frames starting with an intra coded frame.

The MPEG-2 reference software was modified for the compression of the sequences. After motion estimation within the 2D video sequence the vectors were written into a file. Instead of estimating the motion vectors for temporal prediction within the depth map sequence, they were imported from that file. After compression of the depth map sequence by applying the video motion vectors, the coding cost for the vectors in both bit streams may differ. For example by using P-frames the MPEG-2 encoder estimates one motion vector per macro block. If after transformation and quantization all residual coefficients are set to zero and the motion vector is zero no information needs to be transmitted. In that case a macro block skipping is signaled and the decoder will copy the macro block from the previous frame at the specific position. In order to perform that operation no vector is needed and so it is not coded. However the vector predictors are reset. The exact coding cost for the shared motion vectors was determined while coding. Afterwards it was subtracted from the whole MPEG-2 bit stream.

Table 1. Parameter configurations for sequence coding

conf.	description
C_1	frame motion compensation, I- and P-frames
C_2	frame motion compensation, I-, P- and B-frames
C_3	field motion compensation, I- and P-frames
C_4	field motion compensation, I-, P- und B-frames
C_5	adaptive frame/field motion compensation, I-, P- and B-frames

4. EXPERIMENTAL RESULTS

In Fig. 2(a) the method of separate coding of video and depth map sequences is illustrated. B_V labels the amount of data for coding the conventional 2D video, B_D means the same for coding the depth information. B_D is set to 20 percent of B_V to limit the bandwidth overhead as given in the ATTEST project [6]. The overall amount of data B_O is given by

$$B_O = B_V + B_D = B_V + \frac{B_V}{5} = \frac{6}{5} B_V \,. \qquad (2)$$

If the B_O is kept constant and the motion vectors are shared B_D can be used for intra and residual coding completely as shown in Fig. 2(b).

In the following paragraph the results for coding the video sequence at the bitrates $B_{V_1} = 3000$ and $B_{V_2} = 5000$ Kbit/s are considered. The bitrates for the depth map sequence results from eqn. (2) to $B_{D_1} = 600$ and $B_{D_2} = 1000$ Kbit/s. In Table 2 the reconstruction qualities Q_V and Q_D depending on the parameter constellation are listed for the two 3D test sequences *Interview* and *Orbi*. Q_D^* labels the quality of reconstructed depth maps coded with motion vectors estimated separately. A look at Table 2 shows that the approach is applicable at lower bitrates and bidirectional temporal prediction (B-frames). At higher bitrates and parameter configuration C_2 it is more advantageous to code the

Table 2. Objective reconstruction quality (mean PSNR in dB) of 2D video (Q_V) and depth map sequence (Q_D) depending on parameter constellation coded with shared motion vectors at the bitrates $B_{V_1} = 3000$ and $B_{V_2} = 5000$ kbit/s, Q_D^* labels quality of reconstructed depth maps coded with motion vectors estimated separately

Sequence *Interview*						
conf.	Q_{V_1}	Q_{D_1}	$Q_{D_1^*}$	Q_{V_2}	Q_{D_2}	$Q_{D_2^*}$
C_1	40.1	38.7	39.8	41.7	41.4	42.6
C_2	40.5	**37.9**	37.4	42.3	40.6	41.5
C_3	39.0	38.0	39.6	40.7	40.4	42.5
C_4	38.2	**36.7**	-	41.0	**39.3**	37.4
C_5	40.0	**36.9**	-	41.9	**39.7**	37.5
Sequence *Orbi*						
conf.	Q_{V_1}	Q_{D_1}	$Q_{D_1^*}$	Q_{V_2}	Q_{D_2}	$Q_{D_2^*}$
C_1	40.2	36.8	37.3	41.5	38.6	39.4
C_2	40.4	**36.7**	-	41.8	38.6	38.8
C_3	40.1	37.0	37.8	41.5	38.7	40.0
C_4	39.9	**36.7**	-	41.6	**38.6**	-
C_5	40.1	**36.5**	-	41.7	**38.5**	-

depth map sequences using motion vectors estimated separately. A dash in a given table entry indicates the impossibility of coding the sequence at the specified bitrate. In this case the amount of data for motion vector coding exceeds the bit budget, leaving no possibility for coding any frame or residual data.

5. DISCUSSION AND CONCLUSION

This paper described an approach for the compression of 3D video data consisting of 2D video and associated depth map sequence using motion vector sharing. Due to its widespread application the standard MPEG-2 is used for compression of both sequences. In respect of the bitrate limit of 20 percent additional coding cost for the depth map sequence, the proposed method is applicable in the case of bidirectional temporal prediction. If the results are compared with separate motion vector estimation and unidirectional prediction the reconstruction quality of the depth maps is always lower.

There are alternative strategies for the compression of the 3D video sequences. One approach is based on using MPEG-2 for compression of the 2D video and on using up-to-date standards for depth map coding. The recent standards MPEG-4 [7] and H.264 [8] achieve a significantly higher efficiency when applied to the depth map sequences [9]. Further strategies on the basis of different motion compensation and coding concepts are discussed in [10].

6. ACKNOWLEDGEMENTS

This work was accomplished within the Postgraduate Research Program supported by the German Research Foundation (DFG). The authors would like to thank the ATTEST project partners who provided the 3D video test sequences.

7. REFERENCES

[1] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, and I. Sexton, "An Evolutionary and Optimised Approach on 3D-TV," in *Proceedings of International Broadcast Conference '02*, Amsterdam, Netherlands, Sept. 2002, pp. 357–365.

[2] C. Fehn, "A 3D-TV Approach Using Depth-Image Based Rendering (DIBR)," in *Proceedings of Conference in Visualization, Imaging and Image Processing '03*, Benalmadena, Spain, Sept. 2003.

[3] ISO/IEC 13818-2, "Information technology – generic coding of moving pictures ans associated audio information: Video," 1995.

[4] C. Fehn, K. Schüür, I. Feldmann, P. Kauff, and A. Smolic, "Distribution of ATTEST test sequences for EE4 in MPEG 3DAV," in *MPEG02/M9219, ISO/IEC JTC1/SC29/WG11*, Awaji Island, Japan, Dec. 2002.

[5] MPEG Software Simulation Group, "MPEG-2 TM5," `www.mpeg.org/MSSG`, July 1996, Version 1.2.

[6] M. Op de Beeck, E. Fert, C. Fehn, and P. Kauff, "Broadcast Requirements on 3D Video Coding," Cheju, Mar. 2002, ISO/IEC JTC1/SC29/WG11, MPEG02/M8040.

[7] ISO/IEC 14996-2, "Information technology – coding of audio-visual objects – part 2: Visual," 1998.

[8] ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC, "Draft ITU-T Recommendation and Final Draft International Standardof Joint Video Specification," Mar. 2003.

[9] C. Fehn, K. Schüür, P. Kauff, and A. Smolic, "Coding Results for EE4 in MPEG 3DAV," Pattaya, Mar. 2003, ISO/IEC JTC1/SC29/WG11, MPEG02/M9561.

[10] Stefan Grewatsch and Erika Müller, "Evaluation of Motion Compensation and Coding Strategies for Compression of Depth Map Sequences," Accepted for 49th SPIE's Annual Meeting, Denver, August 2004.