

# BIT-RATE ADAPTIVE DOWNSAMPLING FOR THE CODING OF MULTI-VIEW VIDEO WITH DEPTH INFORMATION

*Erhan Ekmekcioglu, Stewart T. Worrall, Ahmet M. Kondoz*

Centre for Communication Systems Research, University of Surrey,  
GU2 7XH Guildford, Surrey, United Kingdom  
{E.Ekmekcioglu, S.Worrall, A.Kondoz}@surrey.ac.uk

## ABSTRACT

In this paper, the potential for improving the compression efficiency of multi-view video coding with depth information is explored. The proposed technique uses downsampling prior to encoding, for arbitrary views and depth maps. A bit-rate adaptive downscaling-ratio decision approach is proposed for certain views and depth maps prior to encoding. Colour and depth videos are considered separately due to their different characteristics and effects on synthesized free view-point videos. The inter-view references, if present, are downsampled to the same resolution as the input video to be coded. The results for several multi-view with depth sequences indicate that using bit-rate adaptive mixed spatial resolution coding for both views and depth maps can achieve savings in bit-rate, compared to full resolution and fixed depth-to-colour ratio multi-view coding when the quality of synthesized view-points are considered. The computational complexity in the encoder is significantly reduced at the same time, since the number of blocks coded is reduced, and hence the number of block mode decisions carried out is reduced.

**Index Terms**— Multi-view Video Coding, multi-view plus depth, mixed resolution coding.

## 1. INTRODUCTION

The issue of MVC has grown in significance in order to make the applications like 3D TV [1] and free view-point TV (FTV) [2] possible. To maintain a wide range of viewing angle in free view-point videos systems, it is necessary to capture and transmit multiple representations of a certain video from arbitrary viewpoints and also their associated per-pixel depth information to assist generating intermediate virtual camera locations. According to [3], coding of depth information should take place in encoder side, particularly for free viewpoint video applications, which brings additional bit-rate load. Multiple view plus depth concept [4][5] is essential in considering depth and colour videos jointly in the sense of coding.

In most multi view plus depth coding applications, the bit-rate of depth maps is restricted to a certain percentage of the bit-rate of colour videos. The higher this rate means the better the quality of a generated arbitrary free viewpoint. However, since the reconstructed colour videos and depth maps have effects on the quality of the final rendered free viewpoint to different extents, it is not always optimal to fix the depth-to-colour bit-rate ratio to a certain percentage. In [6], the authors propose a joint depth/texture bit allocation algorithm which selects a group of quantization parameters for colour and the depth sequences, yielding the highest rendered free viewpoint quality, rather than coding depth sequences at a fixed percentage of the colour videos' rate (in general 10%-15%).

In this paper, we encode both colour videos and depth maps at a wide range of bit-rates after downsampling them with different downscaling ratios, regardless of their inter-view dependencies. In decoding process, they are upsampled back to their original resolution. The idea that downsampling an image to a low resolution, then coding at the lower resolution and subsequently interpolating the result to the original resolution can improve the overall PSNR performance of the compression process [7], motivated us to use such an approach. The technique is based on a trade-off between the distortion introduced by downsampling and distortion introduced by quantization. For a fixed bit budget, increasing the downsampling ratio means that less coarse quantization needs to be used. Thus, more information is lost through downsampling, but less is lost through coarse quantization. Finding the optimum trade-off between the two distortion sources leads to improved compression efficiency. Accordingly, bit-rate savings are achieved at certain qualities for both colour and depth videos, and hence, an overall bit-rate saving is achieved when compared to the case, where there is a fixed ratio between the colour video and depth map bit-rates, regarding the quality of rendered free view point sequences.

Section 2 describes the bit-rate adaptive downsampling scheme separately for colour and depth sequences.

---

This work was developed within VISNET II, a European Network of Excellence (<http://www.visnetnoe.org>), funded under the European Commission IST FP6 programme.

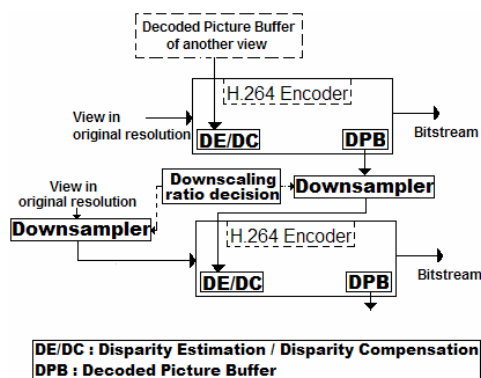
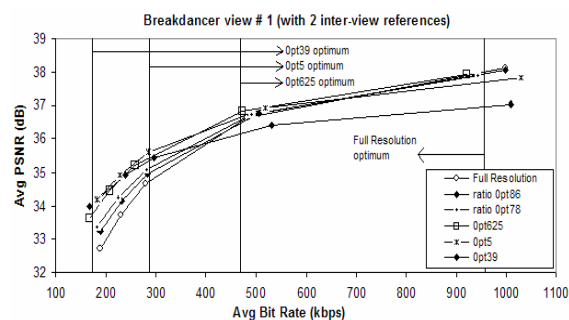


Figure 1. The proposed encoder structure (for a two view case)

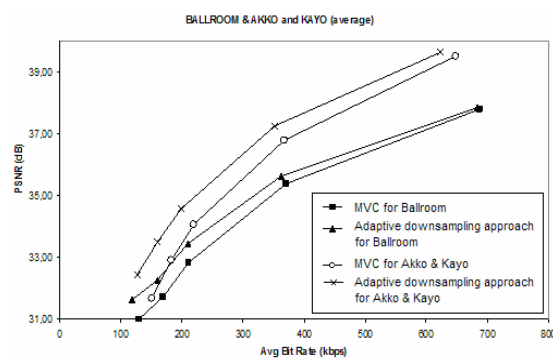
Performance graphs for several test sequences are provided. Section 3 shows the experimental results of view + depth coding with the bit-rate adaptive downsampled coding approach, where the visual comparisons of the results are shown as well. Finally, section 4 concludes the paper.

## 2. BIT-RATE ADAPTIVE DOWNSAMPLING

Figure 1 shows a part of the proposed multi-view encoder structure, for two views using inter-view prediction. As in the draft multi-view coder Joint Multi-view Video Model (JMVM) [8], based on H.264 AVC [9], Hierarchical B frame prediction [10] is used in both temporal and view dimensions. For coding videos (either colour or depth) that use inter-view prediction, the inter-view references are downsampled using the same scaling ratio as the original view before being put in the reference picture buffer, as shown on Figure 1, so as to make sure that all motion and disparity estimation processes takes place between frames of the same spatial resolution. JMVM version 6.0 is used for testing the performance of the proposed method. Each view or depth map is downsampled prior to encoding with a downscaling ratio selected from a set. Divisibility by 16 is maintained for both dimensions of the downsampled videos in order to comply with the codec used. The aspect ratio is kept constant when downsampling. Multi-view videos with provided depth information used in the experiments are of resolution 1024x768 originally (both colour and depth videos). A couple of multi-view videos without any depth information are also tested which are of VGA resolution (640x480) originally. For VGA resolution videos, 0.9, 0.8, 0.7, 0.6, 0.5, 0.3 downscaling ratios are applied to both dimensions of the colour videos and tested separately. Similarly, 0.86, 0.78, 0.625, 0.5 and 0.39 downscaling ratios are applied to both depth and colour videos separately, for the 1024x768 sequences. The difference between the two sets of downscaling ratios arises from the constraints explained earlier in this section. A set of seven filters, designed originally to support the extended range of spatial scaling ratios used in the scalable extension of H.264/AVC, is used to downsample the original views. Integer based 6-



a)



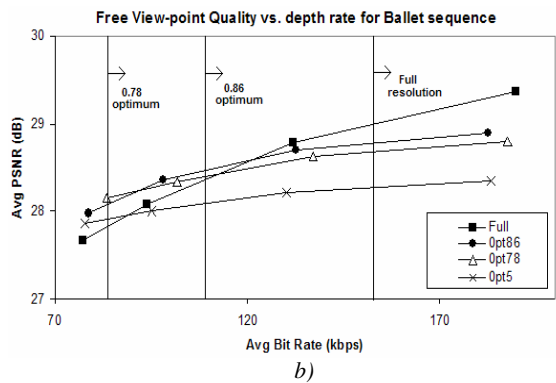
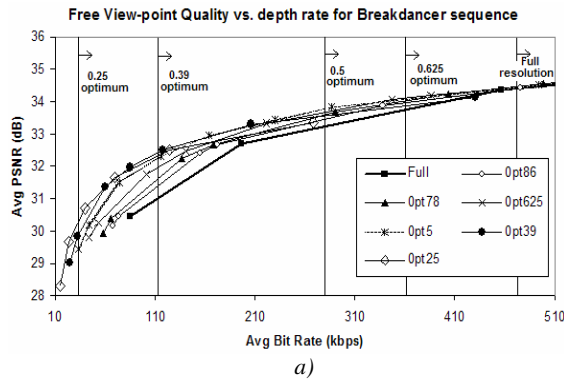
b)

Figure 2. (a) Coding performance of the multi-view coder that uses several downscaling ratios for the second view of the Breakdancer test sequence. (b) Coding performance of the proposed bit-rate adaptive MVC for Ballroom and Akko & Kayo test sequences (average of all views).

tap filters are applied to the decoded low-resolution videos to upsample them back to their original resolutions. 2 multi-view sets with provided depth information and 2 other VGA multi-view sets (colour video only) are used in tests. For all tests, around an operating bit-rate, as the amount of downscaling is increased, the applied quantization step size is decreased, in order to maintain a similar total bit budget.

The results show that different scaling ratios yield different performance characteristics at different bit-rates for both tests. To compute the compression efficiency for the depth maps with different resolutions and quantization parameters, not the reconstruction quality of depth maps is used. Instead, the associated original colour videos and decoded depth maps are used to render a test view on the real camera position closest to the view under concern. The quality of that generated colour video is taken into account. The viewpoint is rendered using depth image based 3D warping algorithm [11]. The results indicate that, regardless of the number of inter-view references used, all videos are affected in the same way.

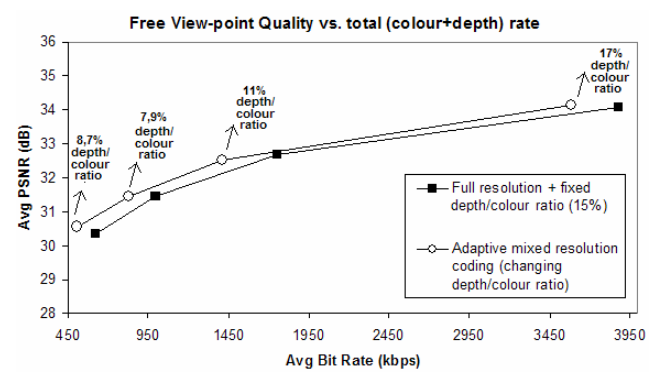
Figure 2 (a) shows the coding performance for the second colour video of the Breakdancer [5] sequence. For the reference method, all sequences are coded in their original resolution, while the other curves show the performance with different scaling ratios. The results show that the



**Figure 3.** (a) Coding performance of the multi-view coder that uses several downscaling ratios for the second view of the *Breakdancer* test sequence. (b) Coding performance of the proposed bit-rate adaptive MVC for *Ballroom* and *Akko & Kayo* test sequences (average of all views).

optimum trade-off between the two distortion types (downscaling and quantization) varies with the target bit rate. The best performance characteristics at medium and low bit-rates, where the quantization distortion is more effective, is achieved with 0.5 - 0.625 scaling ratios, where the coder achieves up to a 28% reduction in bit-rate for the *Breakdancer* sequence. High downscaling ratios, such as 0.39, are only useful at very low bit-rates, where the reconstruction quality is already insufficient to be used in any multi-view application. Reconstruction quality of around 33 dB is satisfactory, while a PSNR of around 35-36 dB is regarded as reasonable quality. Much higher qualities are difficult to distinguish. *Figure 2 (b)* shows the performance characteristics of the proposed technique on a couple of other test sequences (*Ballroom* and *Akko & Kayo*). The results are obtained by taking the average of all views inside a test set. For the proposed technique, 0.6 scaling is used for medium and low bit-rate ranges (less than 400 kbps), and 0.8 scaling is used for high bit-rate ranges (over 400 kbps) at 25 fps.

Similarly, *Figure 3 (a)* and *(b)* show the results for depth map compression performances with *Breakdancer* and *Ballet* sequences respectively. As in the case of colour video



**Figure 4.** Performance comparison of the proposed bit-rate adaptive downsampled coding scheme and full resolution, fixed colour-to-depth rate ratio coding with a depth-to colour ratio of 15% (Total bit-rate is shown including depth and colour bit-rate).

coding, not the full resolution coding yields optimum results at all bit-rates. Especially at medium and low bit-rate points, application of a proper downscaling ratio can make the depth map coding system save up to 55% and 20% of the total depth rate for *Breakdancer* and *Ballet* sequences.

### 3. EXPERIMENTAL RESULTS OF BIT-RATE ADAPTIVE DOWNSAMPLED VIEW AND DEPTH CODER

For a given total bit budget, the bit-rates for depth and colour can be assigned jointly, with the aim of maximizing the rendered free view-point image quality. It is proposed to use the bit-rate adaptive resized colour and depth map coding in achieving this goal, i.e. globally optimizing the total rate-distortion performance, where the bit-rates of both components do not necessarily need to have a fixed ratio (15% in most cases).

The total bit budget is given as:

$$R_{total} = R_{colour}(Q_c, Scale_c) + R_{depth}(Q_d, Scale_d) \quad (1)$$

where,  $R_{colour}$  and  $R_{depth}$  are colour and depth bit-rates,  $Q$  and  $Scale$  are respectively the quantization parameter (Qp) and the downscaling ratio. The optimization problem is defined as:

$$(Q_c^{opt}, Scale_c^{opt}, Q_d^{opt}, Scale_d^{opt}) = \dots \arg \min_{\substack{Q_c \in QSet_c, Q_d \in QSet_d \\ Scale_c \in SSet_c, Scale_d \in SSet_d}} (D_{rendering}(Q_c, Scale_c, Q_d, Scale_d)) \quad (2)$$

under the constraint that:

$$R_{colour}(Q_c^{opt}, Scale_c^{opt}) + R_{depth}(Q_d^{opt}, Scale_d^{opt}) \leq R_{total} \quad (3)$$

where,  $SSet_c$  and  $SSet_d$  are predefined sets for downscaling ratios determined for colour and depth sequences,  $QSet_c$  and  $QSet_d$  represent candidate Qp sets for colour and depth, located around some initial estimate Qp's. Optimizing all

four parameters at the same time requires great computing/coding time and power.

In this paper, a sub-optimal approach is utilized that still outperforms the conventional constant-ratio bit-rate allocation scheme.

For a given total bit budget, 15% is assigned as depth-rate. First, depth map coding is optimized by arranging  $Q_d$  and  $Scale_d$ , such that the necessary bit-rate for depth is decreased optimally, taking into account the rendered free view-point image quality for a constant colour frame rate. Second, optimization is performed for  $Q_c$  and  $Scale_c$ , where the aim is to achieve a reduction in the total colour rate, while maintaining a constant reconstructed colour picture quality. Overall, the total bit-rate needed for a multi-view plus depth map video system has decreased, while the rendered free view-point quality is maintained and depth/colour bit-rate has deviated from the initial assumption of 15%.

Figure 4 shows the coding results for the *Breakdancer* sequence. The result shown is for a single view and its associated depth map (third view and its associated depth map). The free viewpoint video is rendered in the camera position of the second view and the PSNR of the rendered video sequence is evaluated. The PSNR is calculated against the reference which is rendered using the original colour and depth video of the third view. It is clear that there is a total bit-rate save of as much as 33%. Especially, at rendering qualities as low as 30.5 – 31 dB, the depth map is coded at 1/16 of its original size, and the colour video is coded at 39% of its original size, which means that the coding power is also reduced significantly.

Figure 5 visually compares the proposed method with the reference technique at several objective rendering qualities. There is almost no visual difference between the outputs of two methods, although a significant coding rate difference is present.

#### 4. CONCLUSION

In this paper, a joint bit-rate allocation technique for multi-view plus depth coding is proposed that uses bit-rate adaptive spatial sub-sampling. The proposed scheme takes into account the fact that the optimum trade-off between quantization distortion and downsampling distortion varies with operating bit-rate. Experimental results show that the assumption holds for both colour and depth videos. A visual comparison of the proposed method and a fixed rate-ratio method at similar objective qualities shows no difference between the two methods for rendered free viewpoint images. Further investigations concerning the optimization of the downscaling ratio decision process, coding delay and downsampling with different aspect ratios, will be made in the future.

#### 5. REFERENCES

- [1] L. Onural, "Television in 3-D: What are the prospects?", Proceedings of IEEE, Vol. 95, No. 6, June 2007.
- [2] M. Tanimoto, "Overview of free viewpoint television", Signal Processing: Image communication 21, pp. 454-461, 2006.
- [3] M. Tanimoto and et. al., "Proposal on Requirements for FTV", JVT-W127, San Jose, California, USA, April 2007.
- [4] P. Kauff and et. al., "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability", Signal Processing: Image Communication 22, pp. 217-234, February 2007.
- [5] C. L. Zitnick and et. al., "High-quality video view interpolation using a layered representation," ACM Siggraph and ACM Trans. on Graphics Aug. 2004.
- [6] Y. Morvan and et. al., "Joint Depth/Texture Bit-allocation for Multi-view Video Compression", 26<sup>th</sup> Picture Coding Symposium, Portugal, November 2007.
- [7] Bruckstein, A.M.; Elad, M.; Kimmel, R., "Down-scaling for better transform compression," *Image Processing, IEEE Transactions on*, vol.12, no.9, pp. 1132-1144, Sept. 2003.
- [8] A. Vetro, P. Pandit, H. Kimata, and A. Smolic, "Joint Multiview video model JMVM 6.0," ITU-T and ISO/IEC Joint Video Team, Document JVT-Y207, October 2007.
- [9] T. Wiegand, G. J. Sullivan, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T rec H.264 -ISO/IEC 14496-10 AVC)", 7<sup>th</sup> JVT meeting, Thailand, 2003.
- [10] K. Müller and et. al., "Multi-view Video Coding Based on H.264/AVC Using Hierarchical B-Frames", Picture Coding Symposium 2006, China, 2006.
- [11] Sehoon Yea and et. al., "Report on Core Experiment CE3 of Multiview Coding", JVT-T123, Klagenfurt, Austria, July 2006.

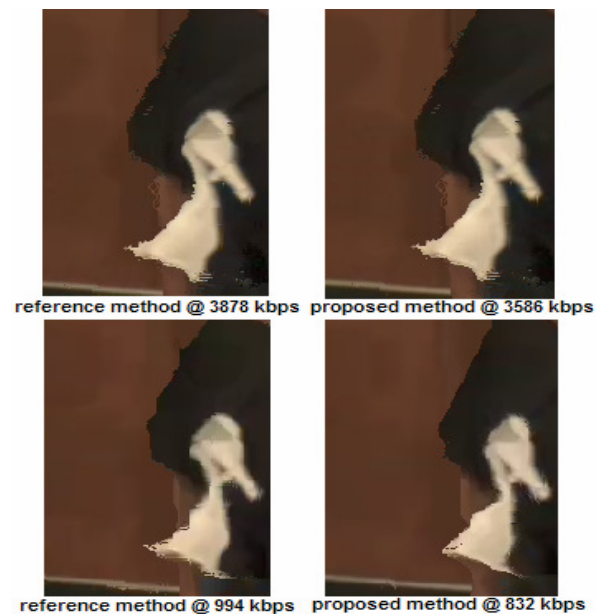


Figure 5. Visual comparison of the proposed technique and the reference technique at 2 different output rendering qualities (results shown for a cropped section of *Breakdancer* test sequence).