3D VIDEO COMPRESSION USING DEPTH MAP PROPAGATION

Sergey Matyunin, Dmitriy Vatolin
smatyunin@graphics.cs.msu.ru, dmitriy@graphics.cs.msu.ru

Department of Computational Mathematics and Cybernetics
Lomonosov Moscow State University
Moscow, Russia

ABSTRACT

We propose a method of 3D video compression based on 2D+depth representation. Only low-resolution depth key frames are transmitted with 2D video. Full-resolution depth map is restored on the decoder's side using 2D video. We evaluated the influence of key frames' resolution, compression ratio and density on the performance of the algorithm. The proposed technique was compared to depth map compression using H.264.

Index Terms **—** Stereo image processing, video compression, three dimensional TV

1. INTRODUCTION

This paper addresses the problem of stereoscopic video compression. Increasing demands to stereo content quality cause the need of the effective video compression algorithms development. The main cue used in stereo video compression is similarity of different views (2 or more) [1]. One of the common approaches to stereoscopic video compression (S3D video) is 2D+depth format usage [2],[3]. This representation can't be used for correct processing of transparent objects and areas behind the objects. Nevertheless this representation is widely used in 2D-to-3D video conversion and as an internal format in a variety of TV's and monitors. 2D+depth format allows generation of arbitrary number of views in a displaying device.

The simplest approach to the depth maps compression is using the traditional video codecs [4]. An important aspect here is determining a ratio between 2D and depth map bitrates that maximizes the resulting quality. Methods of depth maps compression using compressed sensing are developed [3]. Independent compression of depth maps using codecs developed for common video compression is ineffective. 2D video channel information can be used for depth map decoding and restoration. In [5] 2D image is used to increase frame rate and resolution of the depth map obtained using depth sensor. Modified cross-bilateral filtering is used to increase spatial resolution. Frame rate is increased using temporal interpolation. The interpolation is based on motion vectors estimated from 2D video. In [6] joint compression of video and depth map is done using motion vectors estimated from the source video. Additionally, motion in the third dimension is estimated and the obtained motion vectors are transmitted with the compressed video stream. Another important aspect of the 3D video compression algorithms is quality measurement. For various reasons there is no generally accepted method for quality measurement. Research in this direction is in progress [7].

2. PROPOSED SCHEME OF COMPRESSION

The proposed algorithm uses 2D video and the corresponding depth map as input data. In our current research we don't consider 2D video compression. In all experiments we use uncompressed 2D video. Processing pipeline for depth map consists of the following steps:

1) Key frames are selected from input depth map at constant or variable intervals.
   a. Constant intervals are 10, 20, 40 and 100 frames.
   b. Variable intervals are selected adaptively to maximize the quality.
2) Key frames of input depth map are downsampled by constant factor $k = 1$ (without downsampling), 2 or 4.
3) Downsampled key frames are encoded using JPEG 2000 with a constant quality parameter q (64, 128, 265, 512). Size of compressed depth map is measured.
4) Full-resolution key frames are restored using YUVsoft Depth Upscale.
5) Full depth map is restored from key frames using YUVsoft Depth Propagation.

Steps 1–3 correspond to encoding process, and steps 4–5 correspond to depth map decoding. The compression scheme was tested on a set of 9 sequences with different types of motion. Frame rate was assumed equal to 30 fps.

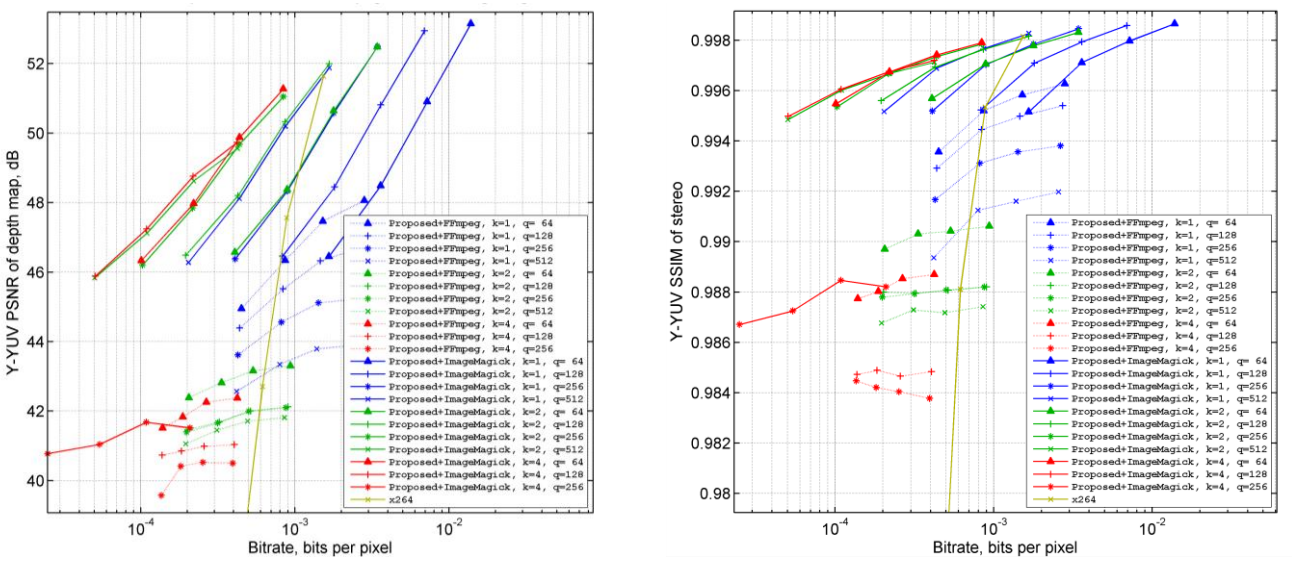
3. QUALITY EVALUATION

The proposed method was compared to the results of depth map compression using x264[1]. We measured the difference in quality for stereo video generated from compressed and original depth map.

Methods of stereo reconstruction from 2D+depth don't provide per-pixel match to the original stereo [7]. Therefore the usage of per-pixel metrics doesn't provide appropriate comparison. Another problem of the 2D+depth representation is occlusion areas filling. Inpainting methods cannot fill large occlusions accurately. Extra information about occlusions should be encoded to provide good quality of restored stereo. We didn't use available original stereo footages in the comparison to eliminate the factors mentioned above. We compared generated stereo for depth map before and after compression (Figure 2a).

In addition we evaluated results using PSNR metric for decompressed depth map (Figure 2b). SSIM metric [8] was used for restored stereo. SSIM metric has higher correlation with human perception therefore we used it for evaluation of stereo quality. SSIM is not suitable for depth map quality assessment because depth map is not a visible decompressed video in the pipeline.

[1] x264 version 0.122.2183.



(a) (b)

Figure 1. Results of the restored stereo quality measurement for the "Bovik2" sequence using a) SSIM and b) PSNR metric. Each line for the proposed method corresponds to the certain set of the spatial resolution decrease and JPEG 2000 compression parameters (k and q respectively) values. Each dot on the proposed method lines corresponds to the certain value of the distance between key frames parameter. After JPEG 2000 codec change restored stereo quality has increased significantly

4. DESCRIPTION OF THE EXPERIMENTS

4.1 Key frames compression using JPEG 2000

We used two implementations of JPEG 2000 codec from FFmpeg[2] and from ImageMagick[3] packages for key frames compression. Codec from ImageMagick demonstrated better compression results both for PSNR and SSIM metrics (Figure 1).

Compression with JPEG 2000 from ImageMagick yields low quality compression on high compression rates q and scale factor k. In this case codec gives single-colour grey image. It is displayed on the graphs by significant quality loss for $k = 4$, $q = 256$ and $q = 512$.

4.2 Disabling psychovisual optimizations in x264

Initially we used the default x264 codec settings. Very slow preset was selected for better compression (command line option "--preset veryslow"). For the second version of the measurements we also disabled optimizations that worsen PSNR and SSIM metric. Psychovisual optimizations are intended for preserving high frequency textures which are not typical for depth maps. It can degrade quality of final stereo video. So to maximize the quality of stereo in SSIM metric (Figure 3) we disabled these options (command line option "--no-psy").

4.3 Increase of key frames density

We analyzed a relation between restored stereo quality and key frames density. Decreasing the distance between key frames down to encoding without depth propagation gives no gain over x264 (Figure 3) due to independent depth map compression for key frames without motion compensation and interframe prediction. Higher key frames density yields SSIM increment, but quality gain is insignificant in comparison with bitrate growth.

4.4 Adaptive key frames placing

Initially key frames in the proposed method were selected at regular intervals. Obviously it is not an optimal strategy. We tested the approach of adaptive key frames selection. Static scenes require less key frames and dynamic scenes must be encoded using dense key frames to achieve acceptable quality.

As initial approximation we took a set of sparse equidistant key frames at the best parameter set (high scale factor, medium JPEG 2000 quality). Then key frames were added semi-automatically to maximize the metric. Example of per-frame SSIM of stereo for a part of "Bovik1" sequence is presented in Figure 4. We tested automatic key frames placement to the minima of SSIM metric and to weighted minima. We didn't get significant compression enhancement and therefore we chose user-guided selection. More complex automatic criteria should be used to deal with complex scenes.

In addition we implemented P-frames analogue in our compression algorithm. While inserting new key frame between existing two we encoded only difference between previous propagation results and original depth map. Sometimes difference contains more high frequencies and requires more space than normal I-frame on the same position. In that case I-frame was inserted instead of P-frame. Example of results for adaptive key frames selection is presented in Figure 5. In this case we optimize SSIM for restored stereo. Visualization of the results is presented in Figure 6, Figure 7.

[2] FFmpeg version N-36088-gdd1fb65 built on Dec 22 2011 12:42:06 with gcc 4.6.2.

[3] ImageMagick version 6.7.6-1 2012-03-14 Q16.

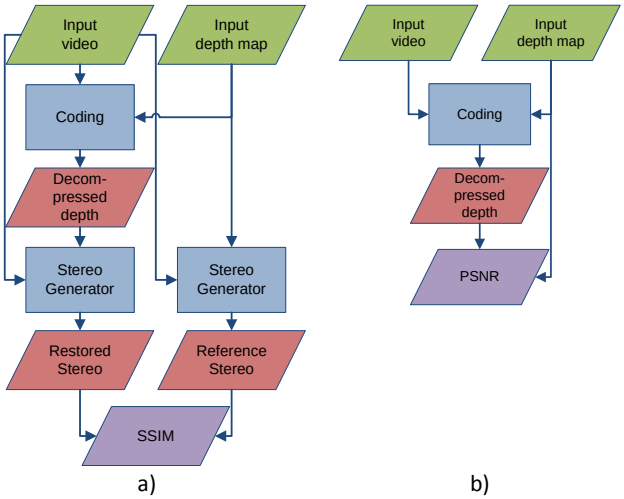


Figure 2. a) The main scheme of the decoded depth map quality measurement. Stereo images restored using the original and the compressed depth maps are compared using SSIM metric. b) Additional depth map quality measurement scheme. Decoded depth map is compared with the original using PSNR metric

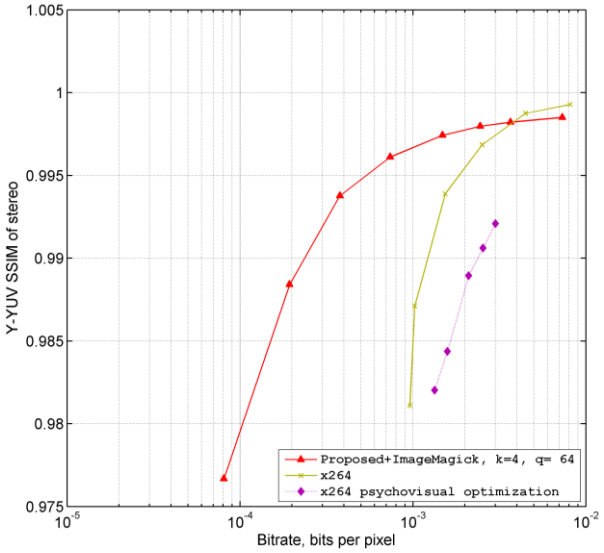


Figure 3. Results for "Pirates" sequence. Performance gain of proposed method decreases with increasing key frames density. Rightmost point of the red graph corresponds to compression without depth map propagation (compression of downsized depth map frames with JPEG 2000)

5. CONCLUSIONS

We proposed a method of depth map compression for multiview video using 2D+depth representation. Only low resolution key frames of depth map are encoded. Depth for the whole video is restored on the decoding stage using information from 2D video. We found out that increased key frames density doesn't provide better compression ratio. Key frames must be selected adaptively to the properties of the depth map and the quality of the restored stereo. We tested several simple automatic approaches to key frames selection, but they were ineffective. We utilized user-guided key frames selection based on SSIM of the resulting stereo. We also implemented inter-frame prediction for intermediate key frames (analogue of P-frames) to improve compression ratio. The described improvements allowed reducing the bitrate up to 50% while preserving the same quality level.

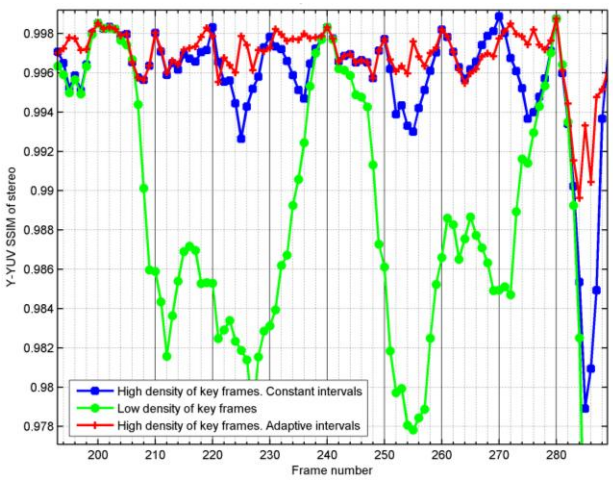


Figure 4. Frame by frame metric values for a part of the "Bovik1" sequence. The blue line corresponds to high key frames density. The green line corresponds to low key frames density. There is a significant decrease of the metric values between the key frames (frames No. 200, 240, 280). Adding of the key frames in the parts with the lowest metric values (red line) allowed to increase quality to the high key frames density configuration level when using lower bitrate

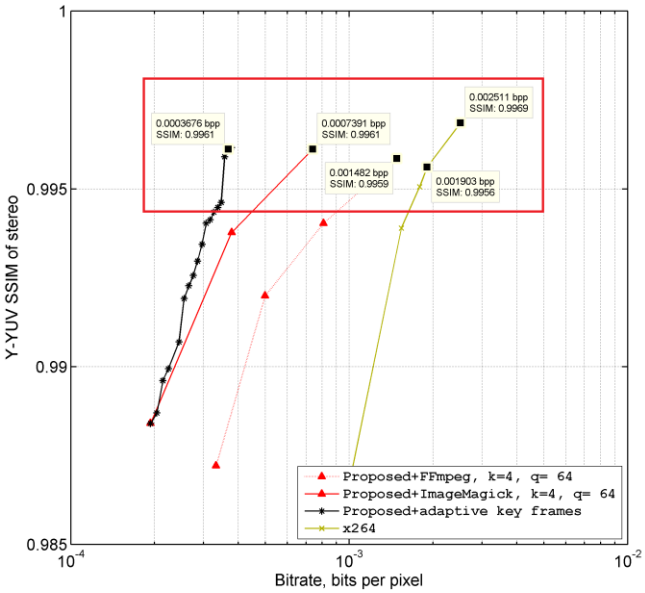


Figure 5. Adaptive key frames distribution reduced the bitrate by 50% on the "Pirates" sequence leaving the SSIM metric values of the same level. The proposed method exceeds the x264 codec results more than 5 times. Each line for the proposed method corresponds to a certain set of the spatial resolution decrease and JPEG 2000 compression parameters (k and q respectively) values. Dots on the x264 line correspond to the different codec compression parameter (crf) values

Further quality improvements must include usage of quadro format to deal with occlusion areas. Uncovered regions must be encoded additionally to prevent artifacts in the areas where Inpainting of stereo generator doesn't provide good quality.

We also intend to implement more complex automatic approaches for key frames selection. They must consider scene changes and confidence metric of depth propagation to select appropriate places for key frames.

In the current part of the work we used constant quality for key frames compression. Another area of further research is optimal selection of key frames compression ratio.

Quality of depth propagation algorithm is very significant for the compression. The major drawbacks of the current version

are forced depth boundaries alignment and depth leakage across objects borders. The first makes difference near objects borders larger if input depth map was not properly aligned to the source video. The second reduces the quality in the dynamic scenes. Eliminating of depth leakage is possible through additional detection and processing of occlusion areas.

6. ACKNOWLEDGEMENTS

This research was partially supported by grant 10-01-00697-a from the Russian Foundation for Basic Research and Intel-Cisco Video Aware Wireless Network Project.

7. REFERENCES

[1] P. Merkle, K. Muller, A. Smolic, T. Wiegand, "Efficient Compression of Multi-View Video Exploiting Inter-View Dependencies Based on H.264/MPEG4-AVC," in *Proc. IEEE International Conference on Multimedia and Expo*, pp. 1717–1720, 2006.

[2] Y. Morvan, P. de With, D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," in *Proc. Stereoscopic Displays and Applications*, SPIE, vol. 6055, pp. 93–100, 2006.

[3] M. Sarkis, K. Diepold, "Depth map compression via compressed sensing," in *Proc. International Conference on Image Processing*, pp. 737–740, 2009.

[4] E. Bosc, V. Jantet, M. Pressigout, L. Morin, C. Guillemot, "Bit-rate allocation for multi-view video plus depth," in *Proc. 3DTV Conference The True Vision Capture Transmission and Display of 3D Video 3DTVCON*, pp. 1–4, 2011.

[5] J. Choi, D. Min, K. Sohn, "2D-plus-depth based resolution and frame-rate up-conversion technique for depth video," *IEEE Transactions on Consumer Electronics*, vol. 56, pp. 2489–2497, 2010.

[6] D. V. S. X. De Silva, W. A. C. Fernando, S. L. P. Yasakethu, "Object based coding of the depth maps for 3D video coding," *IEEE Transactions on Consumer Electronic*, vol. 55, pp. 1699–1706, 2009.

[7] W.-S. Kim, A. Ortega, P. Lai, D. Tian, C. Gomila. "Depth map distortion analysis for view rendering and depth coding," in *Proc. International Conference on Image Processing*, pp. 721–724, 2009.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Figure 6. Fragments of restored left view and SSIM metric visualization for 16th frame of the "Statue" sequence with constant key frames distribution. On the metric visualization lighter areas correspond to lower metric values. SSIM = 0.993088

Figure 7. Fragments of restored left view and SSIM metric visualization for 16th frame of the "Statue" sequence with adaptive key frames distribution. Adaptive key frame choosing allows decreasing amount of the artifacts in the dynamic scenes when compressing with the same bitrate. SSIM = 0.994836