VIEW SYNTHESIS FOR MULTIVIEW VIDEO TRANSMISSION

M. Domański, M. Gotfryd, and K. Wegner
Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology, Poznań, Poland

Abstract - *The paper deals with prospective 3D video transmission systems that are needed for future video services like free viewpoint television or stereovision with autostreoscopic displays. Such systems possible would use transmission of a limited set of viewpoint video sequences. In a receiver, other necessary views have to be synthesized. The paper presents a hybrid technique for viewpoint video synthesis. The original idea is to synthesize a new virtual view from each of reference views separately and then merge them all into one view. In that way the problem of holes in synthesized views is omitted. Moreover, the paper presents the results of measurements (objective and subjective) of quality of synthesized video.*

Keywords: View synthesis, Free viewpoint television, 3D video.

1 Introduction

Prospective 3D video applications will include:

- free-viewpoint television where a viewer is able to control his/her virtual viewpoint and is able to change it freely,
- stereovision with autostereoscopic displays that do not need any annoying glasses to watch video [1].

The abovementioned applications need several views to be available at a receiver. These views may be needed simultaneously for autostereoscopic displays or may be freely selectable as in user navigation in a scene. In both cases, such a requirement would lead to impractically large number of viewpoint video sequences transmitted in the system. Therefore, prospective system would rather transmit very limited number of viewpoint video sequences and the other viewpoint video sequences will be synthesized at the receiver (see Fig.1).

Therefore view synthesis of great research interest recently. In particular, interesting are methods for video synthesis that potentially could be implemented in real time.

Video-based view synthesis rendering in real time is still an open research problem that gains a lot of attention. Recently, many solutions have been proposed [2-5] that often have problems in terms of both computation time and perceptual quality of synthesized views.

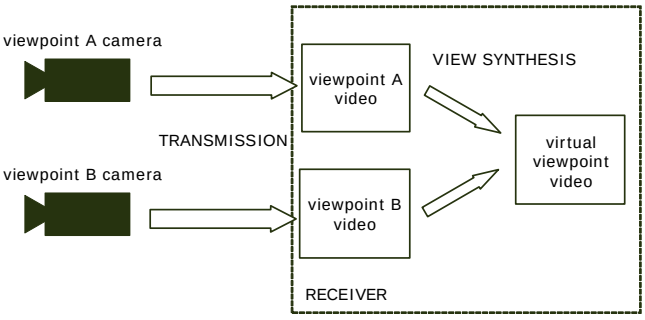


Figure 1. View synthesis for multiview video.

In this paper, described is an efficient and robust technique for view synthesis that has been developed in the course of research related to MPEG standardization activities in Chair of Multimedia Telecommunications and Microelectronics at Poznań University of Technology [6].

2 View synthesis algorithm

View synthesis may use single reference video from a real camera and the respective depth map. Unfortunately, such an approach would suffer from occlusion because the virtual view comprises some regions that are invisible in a single reference view. Therefore virtual video may be synthesized much more correctly from two reference views. Of course, the reference views must be from both sides of the virtual view. Here, we use two reference views together with their depth maps.

In this paper, our main idea is to synthesize a new virtual view from each of reference views separately and then merge them all into one (Fig. 2). Therefore our algorithm is composed of two identical paths. Each of them is aimed at synthesis of the virtual view from one reference view.

Single path algorithm is the following. Position, rotation and other parameters of views are described by the projection

matrices. At the beginning, homography matrices, determining relation between point coordinates in the reference and the virtual view, are calculated, based on their projection matrices. The depth map of a new virtual view is created based on previously calculated homography matrix and depth map of reference view. Then a virtual view is created based on color information from reference view by using depth map and inverse homography matrix. Contour correction stage reduces "ghosting" effect on the edges of objects in scene. In this way single virtual view image is generated with holes in it.

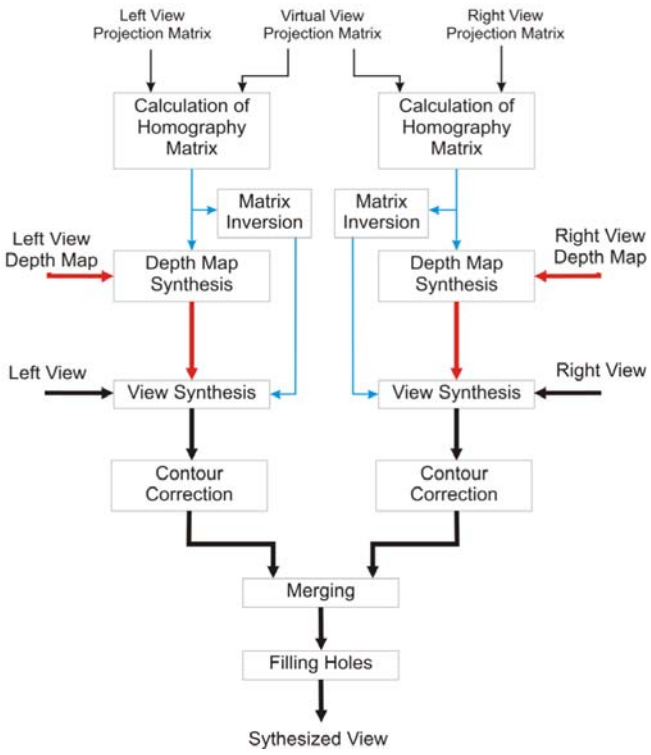


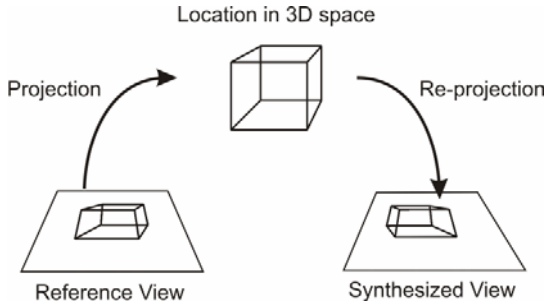
Figure 2. Block diagram of the algorithm proposed.

In the end of each of both parallel path, a virtual view is synthesized. Images from all reference views (paths) are merged together. Unknown regions originating from first reference view are filled with information from second reference view. Nevertheless the virtual view may still have some regions that are contained in none of reference views. These areas must be filled and it is done in filling holes state.

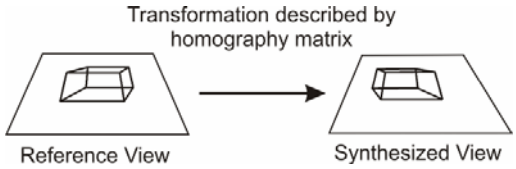
In the paper, we use only two reference views but our algorithm may be simply extended to use as many views as necessary, simply by adding a processing path for supplementary reference view before merge state.

2.1 Calculation of homography matrix

Classic synthesis of new virtual view consists of two steps. Firstly, point coordinates from reference view are projected into its proper location in 3D space. Then 3D points are mapped onto virtual view plane. Such an approach is computationally expensive. Therefore, instead of using projection and re-projection scheme for each point, we use simple view transformation described by homography matrix that defines 2D transformation of one plane into another one. That is the reason why each depth plane in reference view must have its own homography matrix. Since we have 256 possible levels of depth we have to precalculate 256 homography matrixes for each reference view.



a) full transformation through 3D space



b) simple 2D to 2D transformation

Figure 3. Different approach to transformation one 2D view into another.

2.2 Estimation of virtual view depth map

Coordinates of each point in a reference view have been transformed into coordinates of point in the synthesized virtual view by appropriate homography matrix chosen by its depth value. Depth value of each point is stored in depth map buffer of synthesized virtual view. In the case when two points from reference view are transformed into the virtual view, have the same coordinates, always the closer to the camera location (greater depth value) is chosen, because close objects occlude further. Regions that are absent in reference view, such as occluded regions, are marked black in depth map of virtual view and they are considered as unknown. Resultant depth map (of synthesized virtual view) has many small black holes on surfaces which have been rotated during the transformation from reference view into a virtual view

(Fig. 4a). To eliminate those small holes, we use median filtering. The result is shown in Fig. 4b.

a)

b)

Figure 4. Depth map of synthesized view: a) before and b) after median filtering. Brighter points are closer. Black regions have unknown depth.

2.3 Virtual view synthesis

For each point in virtual view the color information is needed. Given created depth maps of virtual view and appropriate inverse homography matrix, we find relation between point coordinates in virtual and reference views. Then the color information is sampled. We use bilinear interpolation to sample appropriate point form reference view. In the end we have several images of synthesized view, one for each reference view.

In order to create one final synthesized virtual view, two synthesized views from two paths are merged into one. Firstly synthesized view from left path is copied to output buffer and then unknown regions are filled with information from second path. The final synthesized view is shown in Fig 6.

a) Left virtual view

b) Right virtual view

Figure 5. Synthesized virtual view from left and right reference view with unknown uncovered regions (black).

2.4 Final rendering

Because of using two reference views, the resultant virtual view still contains only small unknown regions which are occluded in both reference views. Missing areas are interpolated from neighboring pixels as described in [15]. The result of filling is shown in Fig. 6b.

Quality of synthesized view shown on Fig. 6b is satisfactory, but in zoom (Fig. 7a) we can see contour around the uncovered regions replaced with content from another view. Aliasing and blurring on the edges of the object in scene are main reasons of this effect. Our approach to eliminate that effect is simple, but very efficient. In order to eliminate artificial contours, uncovered unknown regions (Fig 5, black points) are outlined by 1 pixel-width. In this way more information will be copied from second synthesized view. Regions extended in such a way can be processed without changes as mentioned before.

a)

b)

Figure 6. Final synthesized virtual view.(a) with unknown regions (black),(b) after filling unknown regions

Figure 7. Magnification of final synthesized view (left) before contour correction (right) after contour correction.

3 Experimental results

The aim of the experiment was to assess quality of synthesized video. The synthesized video is compared to video acquired from a camera. Three standard multiview test video sequences have been used: “Book Arrival”, “Leaving Laptop”, “Alt Moabit” [4]. Each sequence consists of 100 images captured from 16 cameras positioned side-by-side along a straight line with about 65 mm horizontal spacing. Only 3 views (2nd, 3rd and 4th) have been used in the experiments. The 2nd and 4th views have been used as a reference view to synthesized virtual view in-between of them, and 3rd view was used in order to measure quality of the synthesized view.

In order to calculate required depth maps for each reference view, software described in [9] has been used, and depth maps of satisfactory quality have been received. For the sake of simplicity, this software will be called Nagoya software as it has been developed mainly in Nagoya University, Tanimoto Laboratory. Exemplary disparity maps for single frame are shown in Fig. 8. Obtained disparity maps were transformed into the depth maps according to [11].

Figure 8. Exemplary depth maps.

The proposed method has been compared to two other view synthesis methods: the first one was provided by Nagoya University [9] (Nagoya) and the second one provided by Poznań University of Technology [8] (Poznań). The comparison concerns the quality of synthesized images, both whole sequences and each single frame. We have measured objective video quality (by means of PSNR) as well as subjective video quality. In order to estimate the latter one, Mean Opinion Score (MOS) has been measued in a 10-point continuous scale. Rating the quality range from 1 ("very bad with annoying impairments/artifacts") to 10 ("imperceptible"). The reference sequence has been presented before a set of randomly ordered synthesized sequences. In contrast, when comparing single frames, subjects did not know which one was the reference frame. Of course, in both foregoing cases, the order of appearance of synthesized images has been randomly chosen (from between three available view synthesis methods). The test has been carried out on the group of 15 human subjects. Collected opinions have been averaged, with resulting mark given with a 99 % confidence interval. It is worth noticing that sequences are relatively short in duration. Therefore each sequence has been presented twice.

Very good quality of synthesized images is preserved also in case of Nagoya [9] algorithm. The result obtained with the Poznań [2] algorithm differ from two remaining marks in case of "Alt Moabit" sequence because the quality of estimated depth maps for this sequence is worse when compared with the others sequences. Our experiment shows that Poznań [2] algorithm is very sensitive to the quality of depth maps used in synthesis. Two remaining methods are not so vulnerable in this matter and synthesize images of very good quality for "Alt Moabit" sequence. Our results for single frames from "Leaving Laptop" sequence are slightly worse than those from "Alt Moabit", but still received marks are at high level.

It can also be noticed that subjective quality of whole sequences is considered to be worse than single frames taken from the same sequence. In particular, this property holds true in case of "Leaving Laptop" sequence. For Poznań [8] algorithm, perceived subjective quality has increased from 5,11 for whole sequence to value 7,22 in case of single frame (Fig. 10, Fig. 9). The same property can be seen in "Alt Moabit" sequence (Fig. 11, Fig. 9). Fluctuation of depth values in consecutive frames of sequence, which takes place near the edges of static objects in the scene, has negative influence on subjective quality of synthesized sequence. For single frames that effect does not occur, and resulting mark of subjective quality is significantly higher.

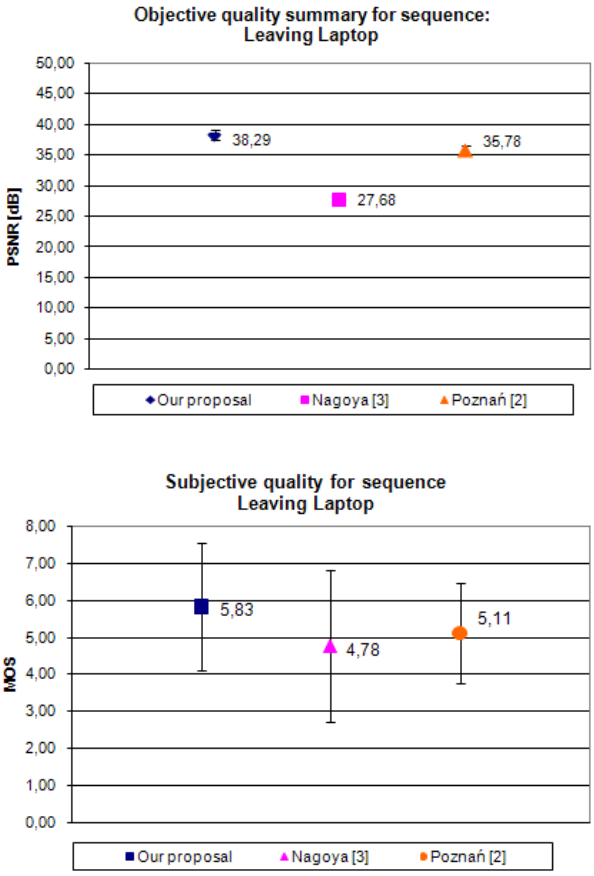


Figure 9. Objective and subjective quality of sequence "Leaving Laptop".

In case of "Alt Moabit" sequence, we can observe that Nagoya [3] algorithm, as well as Poznań [8] algorithm have 12 dB lower PSNR compared to our proposal (Fig. 11). However, obtained PSNR values for Nagoya [9] algorithm do not correspond to the subjective sequence quality perceived by users. This property is proved also in case of "Leaving Laptop" sequence (Fig. 10). Our study shows that for Nagoya technique [9], PSNR measures are not enough correlated inadequate results in comparison with MOS. It has turned out that images synthesized with the usage of Nagoya [9] view synthesis tool are shifted with respect to desired view position. Therefore the results obtained with PSNR measure in case of Nagoya [9] software are inappropriate.

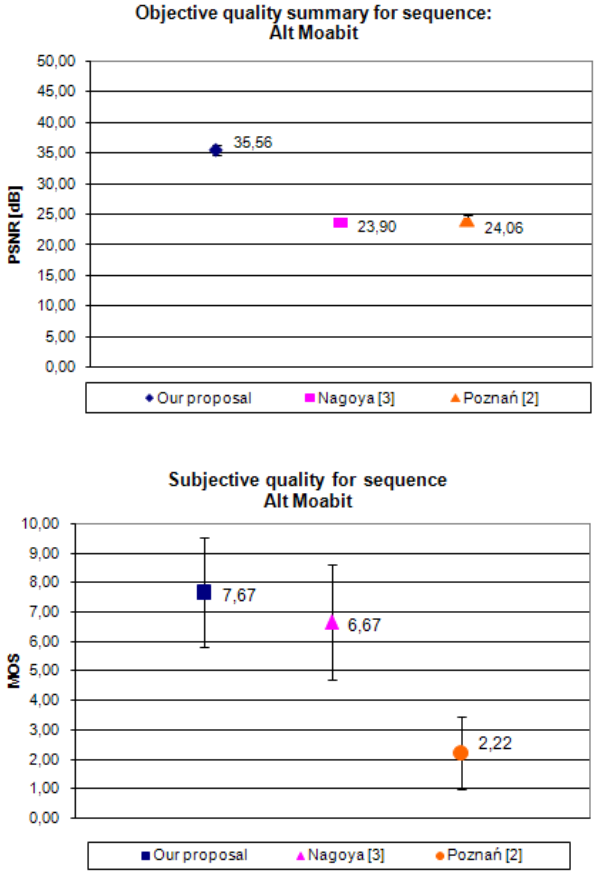


Figure 10. Objective and subjective quality of sequence "Alt Moabit".

Our proposal received the best results with respect to PSNR and also in case of subjective quality measure (Figs. 10, 11). In case of subjective quality of sequences the difference is not significant, but still in favor of our proposal.

4 Conclusions

In this paper we have presented a new fast and robust view synthesis have been described. The experiments proved very high quality of the synthesized video. The technique is implenetable in real time using standard pC computers. The original synthesis improvements have been described together with respective reference software of MPEG.

In the experiments, the synthesis was fully implemented, therefore, the experimental results are quite reliable.

5 Acknowledgements

This work was supported by the public founds as a research project in years 2007-2009. Prof. M. Domański is the recipeinet of Award MISTRZ form Foundation for Polish Science.

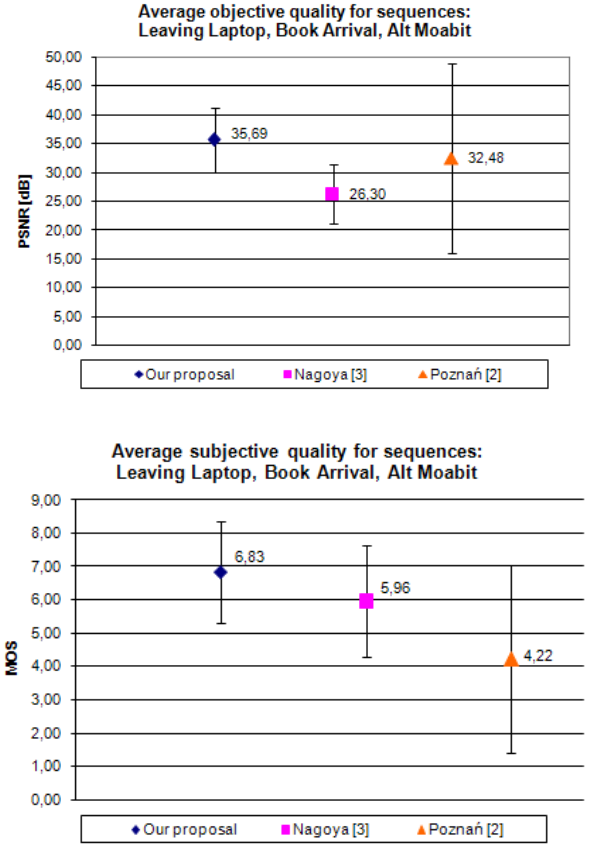


Figure 11. Average objective (left) and subjective (right) quality of sequences "Leaving Laptop", "Book Arrival", "Alt Moabit".

6 References

[1] Jung-Young Son, B. Javidi, Kae-Dal Kwack, "Methods for Displaying Three-Dimensional Images", Proceedings of the IEEE, vol. 94, no. 3, March 2006.

[2] S. Jo, D. Lee, Y. Kim, Ch. Yoo, "Development of a simple viewpoint video system", IEEE Int. Conf. Multimedia and Expo, Hannover, June 2008, pp. 1577-1580.

[3] H. Kimata, S. Shimizu, Y. Kunita, M. Isogai, K. Kamikura, Y. Yashima, "Real-time MVC viewer for free viewpoint navigation", IEEE Int. Conf. Multimedia and Expo, Hannover, June 2008, pp. 1437-1440.

[4] Ch. Yeo, J. Wang, K. Ramchandran, "View synthesis for robust distributed video compression in wireless camera networks", IEEE Int. Conf. Image Processing, ICIP 2007, pp. III-21 – III-24.

[5] S. Yea, A. Vetro, "View synthesis prediction for rate-overhead reduction in FTV", 3DTV Conference, may 2008, pp. 145-148.

[6] M. Gotfryd, K. Wegner, M. Domański, „View synthesis software and assessment of its performance“, ISO/IEC JTC1/SC29/WG11 (MPEG) M15672, Hannover, Germany, July 2008.

[7] O. Stankiewicz, K. Wegner, “Depth Map Estimation Software”, MPEG 2008/M15175, Antalya, Turkey, January 2008.

[8] O. Stankiewicz, K.Wegner, Depth Map Estimation Software version 2, MPEG 2008/M15338, Archamps, France, April 2008

[9] M. Tanimoto, T. Fujii and K. Suzuki, “Multi-view depth map of Rena and Akko & Kayo”, ISO/IEC JTC1/SC29/WG11, M14888, October 2007.

[10] Ingo Feldmann I., Kauff P., Mueller K., Mueller M., Smolic A., Tanger R., Wiegand T., Zilly F., HHI Test Material for 3DVideo, MPEG2008/M15413, Archamps, France, April 2008

[11] “Description of Exploration Experiments in 3D Video”, MPEG 2008/N9596, Antalya, Turkey, January 2008.

[12] M. Pollefeys, R. Koch, L. Van Gool, “A simple and efficient rectification method for general motion”, Proc. International Conference on Computer Vision 1999, pp. 496-501.

[13] “Call for Contributions on FTV Test Material”, ISO/IEC JTC1/SC29/WG11, MPEG 2007/N9468, Shenzhen, China, October 2007.

[14] J. Sun, N.N. Zheng, and H.Y. Shum, “Stereo matching using belief propagation” IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(7):787–800, 2003.

[15] Bertalmio M., Bertozzi A. L., Sapiro G., Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting, Proceedings of the International Conference on Computer Vision and Pattern Recognition , IEEE, Dec. 2001, Kauai, HI, vol. I, pp. 355-362.