INTERNATIONAL ORGANISATION FOR STANDARISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG/M15672
July 2008, Hannover, Germany

Title	**View synthesis software and assessment of its performance**
Sub group	**Video**
Authors	**Mateusz Gotfryd**, **Krzysztof Wegner** (kwegner@multimedia.edu.pl) and **Marek Domański** (domanski@et.put.poznan.pl), Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland

1 Introduction

This document is prepared in response to N9468 "Call for Contributions on FTV Test Material" [7], in particular for "Depth Map Estimation & View Synthesis Software" paragraph. The document presents a hybrid technique and software for video synthesis. Moreover, the document presents the results of measurements (objective and subjective) of quality of synthesized video.

2 View synthesis algorithm

For a virtual camera, view synthesis may use video from a real camera (the real reference video) and the respective depth map. Unfortunately such an approach suffers from occlusion. The virtual view comprises some regions that are invisible in the reference view. Therefore virtual video may be synthesized much more correctly if two reference views are used. Of course, the reference views must be from both sides of the virtual view. In this document we use two reference views and their depth maps.

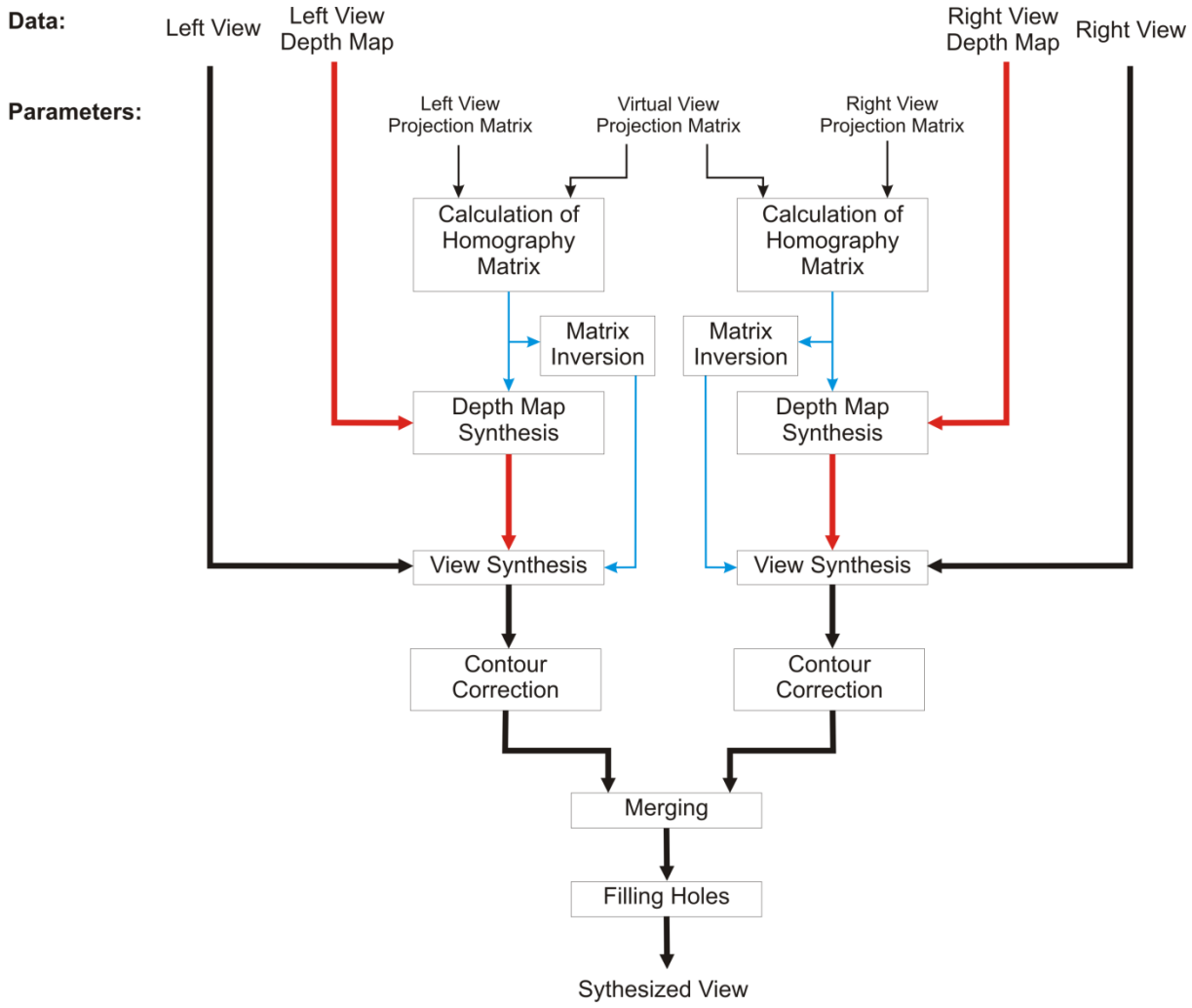

Figure 1. *Block diagram of the algorithm proposed.*

Our main idea is to synthesize a new virtual view from each of reference views separately and then merge them all into one (Fig. 1). 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 is processed as follows. 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.

Both paths of algorithm are parallel. In the end of each path synthesized virtual view is created. 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.

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

Calculation of homography matrix

Usually the 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. This approach requires drastic computational costs. That is why instead of using projection and re-projection scheme for each point we use simple view transformation described by homography matrix. Homography matrix 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.

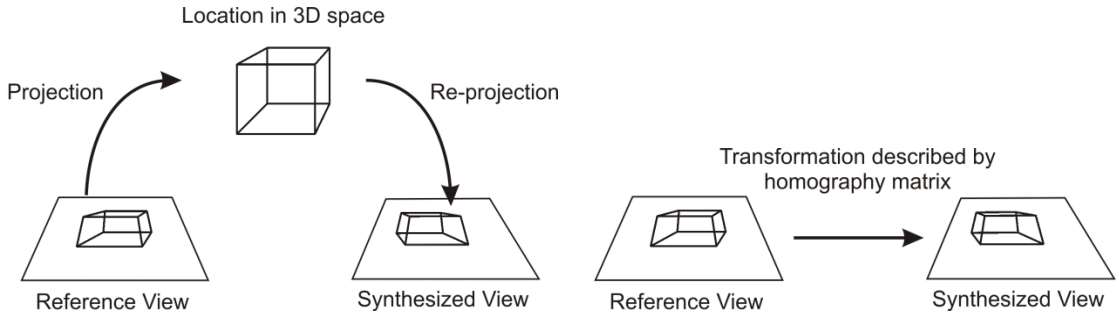

Figure 2. *Different approach to transformation one 2D view into another.*
a) full transformation through 3D space, b) simple 2D to 2D transformation

Creation 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. 3a). To eliminate those small holes, we use median filtering. The result is shown in Fig. 3b.

a) b)

Figure 3. *Depth map of synthesized view: a) before and b) after median filtering.*
Brighter points are closer. Black regions have unknown depth.

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.

a) b)

Figure 4. *Synthesized virtual view from left (a) and right (b) reference view with unknown uncovered regions (black).*

Virtual view image merging

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 5.

a) b)

Figure 5. *Final synthesized virtual view.*
a) with unknown regions (black), b) after filling unknown regions

Filling holes

In spite of using two reference views, resultant virtual view still contains small unknown regions which are occluded in both reference views. Missing areas are interpolated from neighboring pixels as described in [9]. The result of filling is shown in Fig. 5b

Contour correction

Quality of synthesized view shown on Fig. 5b is satisfactory, but in zoom (Fig. 6a) 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 4 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. *Magnification of final synthesized view*
a) before contour correction b) after contour correction

3 Experiment description

The aim of the experiment was to measure quality of the synthesized video. The synthesized video is compared to video acquired from a camera.Three standard test video sequences have been use. These are the sequences provided by Fraunhofer HHI [4]: “Book Arrival”, “Leaving Laptop”, “Alt Moabit”. 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 (2^{nd}, 3^{rd} and 4^{th}) have been used in the experiments. The 2^{nd} and 4^{th} views have been used as a reference view to synthesized virtual view in-between of them, and 3^{rd} view was used in order to measure quality of the synthesized view. The idea is shown in Fig. 7.

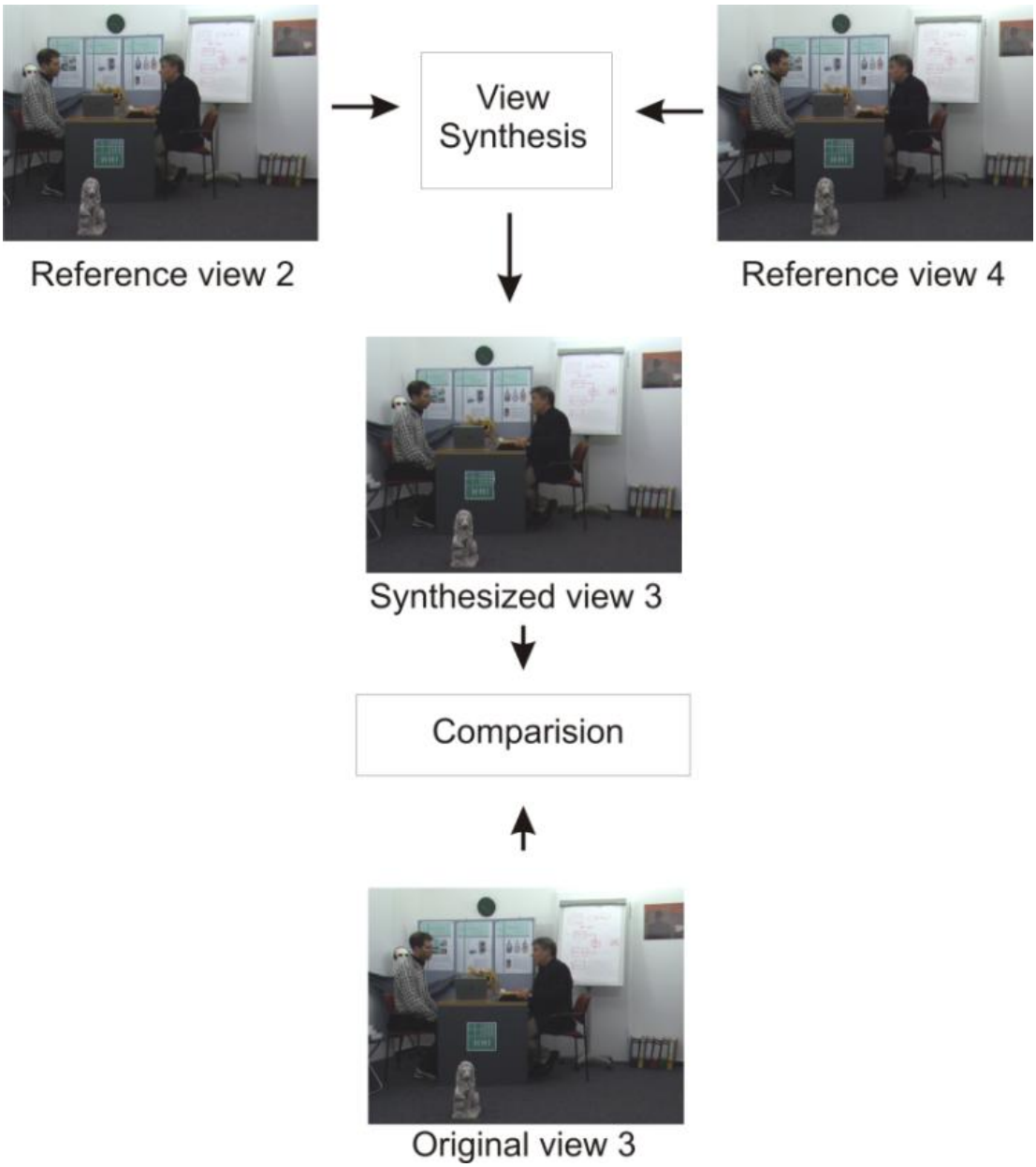

Figure 7. *The idea of a system used for measurement of the quality of view synthesized in-between two reference views with their disparity maps.*

In order to calculate required depth maps for each reference view, software described in [3] has been used, and depth maps of satisfactory quality have been received. Exemplary disparity maps for single frame are shown in Fig. 8. Obtained disparity maps were transformed into the depth maps according to [5].

Figure 8. *Exemplary frames (top) with depth maps (bottom) from test sequences (from left): "Book Arrival", "Leaving Laptop", "Alt Moabit".*

The proposed method has been compared with two other view synthesis methods: first one provided by Nagoya University [3] (Nagoya) and second one provided by Poznań University of Technology [2] (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. To evaluate the second one, we have used the idea of Mean Opinion Score (MOS). The MOS (in case of our study) is expressed by 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.

4 Results

As shown in Fig. 9, a single frame synthesized with proposed algorithm is perceived almost as undistinguishable from reference frame in case of both sequences. Very good quality of synthesized images is preserved also in case of Nagoya [3] 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.

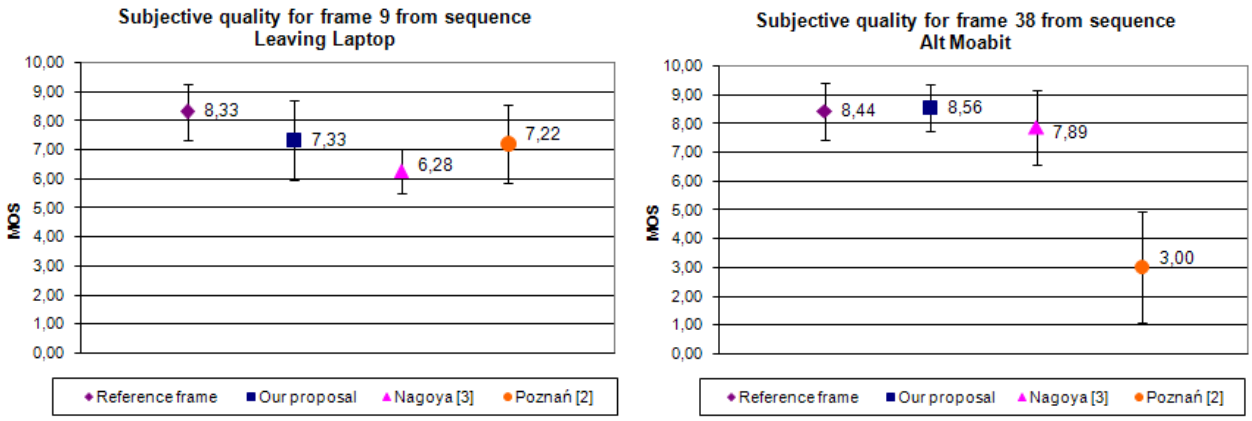

Figure 9. *Subjective quality of single frame from sequence "Leaving Laptop" and "Alt Moabit".*

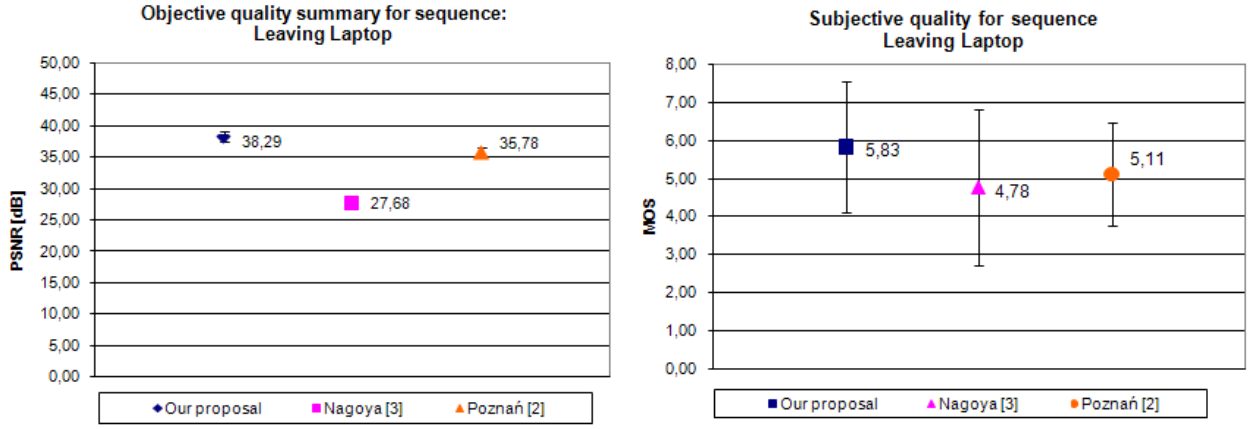

Figure 10. *Objective and subjective quality of sequence "Leaving Laptop".*

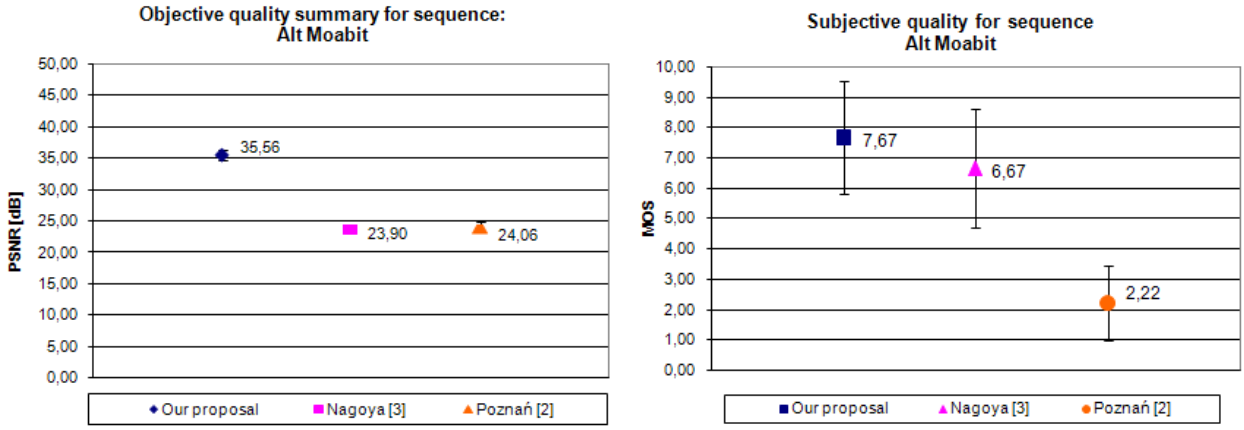

Figure 11. *Objective and subjective quality of sequence "Alt Moabit".*

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 [2] 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.

In case of "Alt Moabit" sequence, we can observe that Nagoya [3] algorithm, as well as Poznań [2] algorithm have 12 dB lower PSNR compared to our proposal (Fig. 11). However, obtained PSNR values for Nagoya [3] 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 PSNR measure concerning Nagoya [3] gives inadequate results in comparison with MOS. It has turned out that images synthesized with the usage of Nagoya [3] view synthesis tool are shifted with respect to desired view position. Therefore the results obtained with PSNR measure in case of Nagoya [3] software are inappropriate.

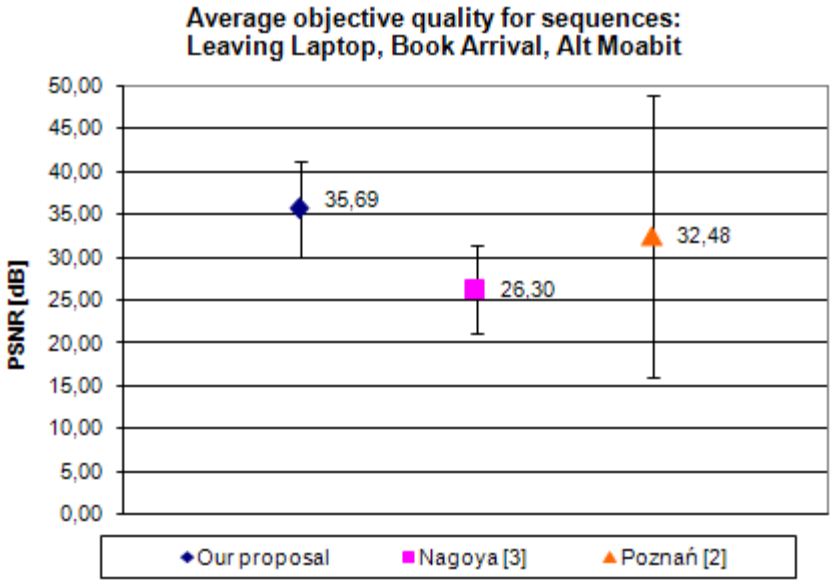

Figure 12. *Average objective quality of sequences "Leaving Laptop", "Book Arrival", "Alt Moabit".*

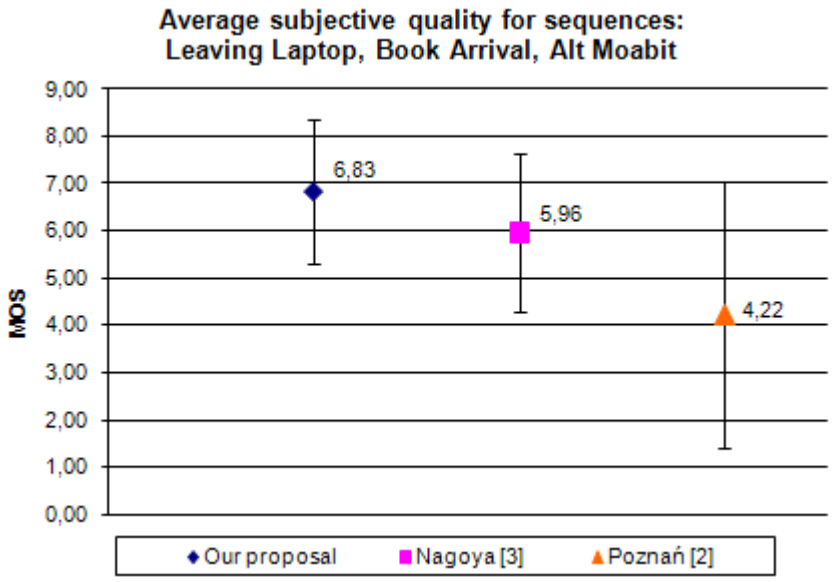

Figure 13. *Average subjective quality of sequences "Leaving Laptop", "Book Arrival", "Alt Moabit".*

Our proposal received the best results with respect to PSNR and also in case of subjective quality measure (Fig. 12, Fig. 13). In case of subjective quality of sequences the difference is not significant, but still in favor of our proposal.

5 Conclusions

In this paper we have presented a new view synthesis software and subjective experiments carried out with view-synthesis software provided to MPEG. The improvements have been described both from technical and conceptual point of view. Our proposal received the best results not only in criteria of PSNR, but also in case of subjective quality measure (MOS) among considered algorithms.

6 Acknowledgements

This work was supported by the public founds as a research project in years 2007-2009.

7 References

[1] O. Stankiewicz, K. Wegner, "Depth Map Estimation Software", MPEG 2008/M15175, Antalya, Turkey, January 2008.

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

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

[4] 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

[5] "Description of Exploration Experiments in 3D Video", MPEG 2008/N9596, Antalya, Turkey, January 2008.

[6] 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.

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

[8] 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.

[9] 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