STEREOSCOPIC DEPTH REFINEMENT BY MID-LEVEL HYPOTHESIS[1]

Olgierd Stankiewicz, Marek Domański, Krzysztof Wegner

Poznań University of Technology,
Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland
Email: [ostank,kwegner]@multimedia.edu.pl

ABSTRACT

In this paper we propose a new technique for refinement of depth maps. The technique exploits a stereoscopic pair of images and full-pixel precision disparity map in order to produce the output disparity map with sub-pixel precision. The technique employs view synthesis for verification of hypotheses on disparity values, which are formed to iteratively improve the disparity map. Results of experiments, presented for still images and video sequences, show that proposed technique increases quality of depth and has low computational cost. Applications of the proposed technique can be found in future real-time 3D video broadcasting systems which require fast and robust disparity estimation algorithm with high precision.

Keywords— Stereoscopic depth, precision refinement, post-processing

1. INTRODUCTION

Modeling of 3D scenes is an important task in many applications related to computer vision, and probably it will be one of the key issues related to future three-dimensional (3D) video systems such as 3D television, 3D video servers and free-view television. Recently, 3D video has gained a lot of attention among research community, especially ISO/IEC has initiated works towards standardization of 3D video representations. It is very likely that future 3D video representations will also comprise information on stereoscopic depth for all pixels. For a picture, such information is called a depth map.

Accuracy of the above mentioned 3D video representation strongly depends on quality of the provided depth map, i.e. its spatial resolution and precision of the depth value at each point of the image.

There are two main approaches that lead to obtaining the depth maps. First approach is to apply a specialized depth camera that incorporates infrared sensors. The currently available depth cameras are not only quite expensive but unfortunately they produce depth maps with relatively low spatial resolution. Moreover, they provide depth measurements in a limited range (typically of about few meters from the camera).

Therefore, another approach gains a lot of attention among research community. The main, well-known idea of that approach is to estimate depth by analysis of images acquired from two or more cameras with parallel optical axes. Such techniques indirectly calculate depth by stimation of disparities between positions of individual scene elements in distinct views. Depth itself is reciprocal to disparity and thus conversions in both directions are unique [1].

This paper focuses on the latter approach. There exist many methods for disparity estimation [2–5]. Modern state-of-the-art disparity estimation techniques comprise optimization using even iterative algorithms like Belief Propagation or Graph Cuts [2,6,7]. These algorithms are robust, but their complexity increases vastly with resolution and precision of requested disparity maps. Nevertheless, precision of disparity estimation is an important issue for 3D video representations. In such applications disparity is estimated with accuracy to fractions of the spatial sampling periods in images, i.e. with sub-pel accuracy. In Section 2, it will be shown that direct sub-pel disparity estimation could be difficult in the context of future real-time applications.

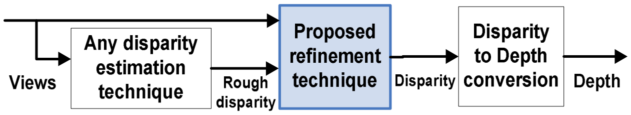

Views
Any disparity estimation technique
Rough disparity
Proposed refinement technique
Disparity
Disparity to Depth conversion
Depth

Fig. 1. Scope of the paper within the depth estimation process.

In order to avoid the above mentioned problems, we use two-step disparity estimation (Fig. 1). In the first step, disparity is roughly estimated. Usually its accuracy is limited to the sampling period, i.e. full-pel accuracy. In the second step the disparity map is refined to sub-pel accuracy. This paper deals with the second step: the refinement of rough disparity maps into maps with accuracy to fractions of the spatial sampling period. It is done in an iterative evaluation-by-hypothesis process.

We have encountered few papers considering similar approaches. In [8], depth map is refined with the use of ego-motion of the camera. This requires stereoscopic video

[1] This work was supported by the public funds as a research project.

sequence with a motion of the camera and thus is not suitable for generic cases of processing. Authors of [9] have proposed an algorithm for depth map improvement with anisotropic diffusion. This method provides smooth, high-precision depth maps, but unfortunately it does not preserve depth discontinuities over the edges of the objects. The idea of evaluation-by-hypothesis, similar to ours, is employed in paper [10]. Mumford-Shah-like functional is employed to compute a smooth depth map, basing on multiple depth hypotheses obtained from different matching algorithms. The certainty of the depth is not considered though, which is a drawback of that proposal.

Regardless of the results mentioned above, currently, there is lack of fast post-processing techniques that could improve the precision of a disparity map and well preserve spatial edges. This observation is a key motivation for our work. In this paper we propose a post-processing algorithm (Fig. 1) that increases precision of disparity maps, preserves spatial edges and is not computationally expensive.

2. PROBLEM ANALYSIS

Most of the current state-of-the-art disparity estimation techniques produce disparity maps in a single-step procedure. The complexity of disparity estimation increases with the number of disparity levels, and unfortunately it increases faster than the growth of benefits from attained higher precision. As our research revealed, increasing the number of disparity levels vastly increases the complexity of disparity estimation, but the fidelity of 3D scene model tends to saturate (Fig. 2). This is a rationale of the statement, that it is not efficient to estimate high-precision depth maps in a regular single-step process. In Fig. 2, typical results are presented for the state-of-the-art technique [6,7] widely used as reference by the ISO expert group MPEG. Quality of the disparity map (and the respective depth map) is measured by the quality of a synthesized view (see Section 5.2 for more detailed explanation). The results of Fig. 2 are calculated for Book_arrival multiview test sequence [14] with spatial resolution of 1024×768. Nevertheless, we have obtained similar results for other test sequences.

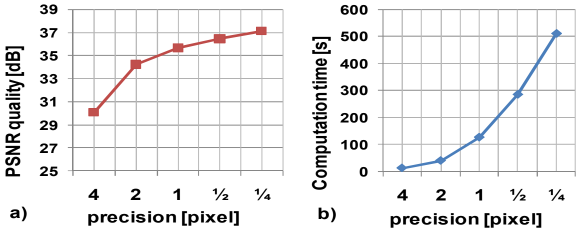


Fig. 2. Quality (a) and computation complexity (b) of disparity estimation as a function of disparity precision.

Therefore, in typical scenario of depth map estimation, disparities are estimated with only full-pel precision. Unfortunately, such accuracy is insufficient for most applications related to 3D video. This lack of precision is especially noticeable in the case of continuous flat surfaces that are nearly (not exactly) perpendicular to the optical axis of the camera (Fig. 3). In the corresponding disparity map, there exists a false contour as a result of full-pel accuracy of disparity. Such a false contour may be observed as an unit-step edge (Fig. 3 left) that results in severe artifacts that can be found in reconstruction of a 3D scene. These artifacts could be substantially reduced by refining the disparity map to sub-pel accuracy. In the case of half-pel refinement, a unit-step false edge would be replaced by two half-step edges (Fig. 3 right). This would significantly yield the reduced artifacts in the reconstructed 3D scene.

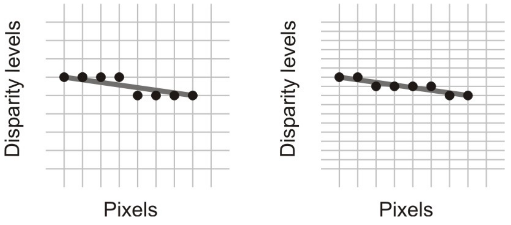


Fig. 3. False contours in disparity maps for a surface being nearly perpendicular to the camera axis left - for full-pel accuracy, right – for half-pel accuracy

In this paper, we propose an original disparity refinement technique called mid-level hypothesis (MLH) technique. This technique identifies the false edges in a rough disparity map. Then, at individual pixels, the technique introduces mid-level (the intermediate level) values of disparity in order to reduce the false contours of disparity maps. In that way the disparity quantization step is halved. Of course, this technique may be used iteratively. After n iterations, the disparity quantization step is reduced by factor 2^n.

3. IDEA OF THE PROPOSED ALGORITHM

At the input, there is a disparity map with limited precision. Such a map may result from any state-of-the art disparity estimation technique. Moreover, at the input, there is a stereoscopic pair of two pictures. The primary view is the view that corresponds to the disparity map processed. The other, side view will be called a reference view.

The above-described set of image and depth data corresponds to a very likely scenario of future transmission of 3D video. The idea of the paper is to exploit this available information in order to improve depth map estimation. Moreover, the proposed approach may be used for efficient transmission of video and depth information. Only rough depth maps need to be transmitted with low bit-rate while the exact depth may be calculated in the receiver using the technique proposed.

Our technique is aimed at improving the accuracy of the disparity map. The idea is to use the a side reference view as additional information for this improvement.

The reference view is used in the following way. From the primary view and the rough disparity map, a synthesis of reference view is obtained. The more accurate is the disparity map, the more the synthesized reference view is similar to the actual reference view. Therefore, similarity of the two versions of the reference view may be used as an indicator of the accuracy of the depth map.

At the beginning of the process, the edges that correspond to the disparity quantization step are identified. For full-pel disparity map, these are unit-step edges. For the sake of simplicity we will use this name herein. The disparity map is processed only locally along those potentially false contours (Fig. 4). The potential improvement may be attained by introducing the mid-level values into the disparity map.

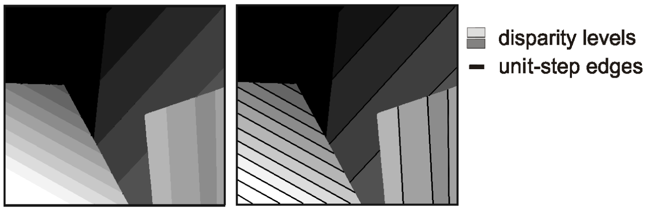


Fig. 4. Disparity map (left) and the same disparity map with marked unit-step edges (right) [3].

False edges occur when an inaccurate rough estimation is forced to quantize disparity levels. The unit-step edges may be false contours or may also represent actual depth differences: there is an uncertainty to be resolved. In our refinement algorithm it is done by verification of the mid-level hypothesis. At first, our algorithm assumes that each pixel neighboring to unit-step edges (Fig. 5) in the disparity map, should have the intermediate disparity level. Next this hypothesis is verified for each pixel.

Together with those potentially false edges, a problem arises that has to be addressed at each individual pixel: Should we change the disparity value to a mid-level value or not? The question is answered by testing the results of the assumed change to mid-level value. After such a change local synthesis of the reference view is done. Then, this portion of the synthesized side view is compared to the same portion of the synthesized view obtained without the disparity change to the mid-level. The hypothesis of the disparity mid-level value is verified if the newly synthesized reference view is more similar to the real reference view as compared to the reference view synthesized using the input rough disparity map.

In principle we assume "mid-level hypothesis" at each edge pixel, i.e. we set the pixel value to intermediate value. Then, we verify the hypothesis by checking if the synthetic size view is more similar to the original reference view.

The mid-level hypothesis is spread from an unit-step edge. Spreading stops when no point passes the verification test (Fig. 6).

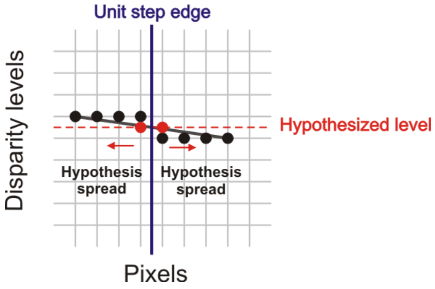


Fig. 5. Spreading of mid-level hypothesis starting from an *unit-step* edge.

Therefore, the precision of the disparity map is improved by insertion of intermediate disparity levels in-between of existing levels. The proposed technique never degrades the processed disparity map.

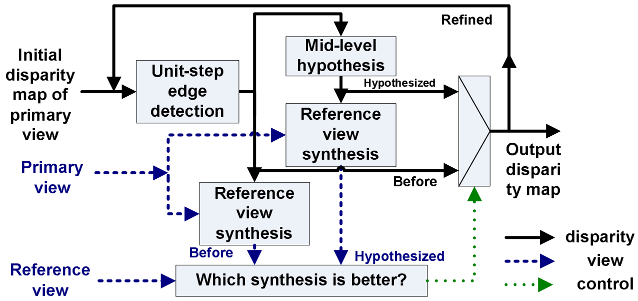


Fig. 6. Scheme of proposed disparity refinement algorithm.

4. IMPLEMENTATION OF THE ALGORITHM

The steps of the algorithm are as follows:

4.1. Detection of unit-step edges.

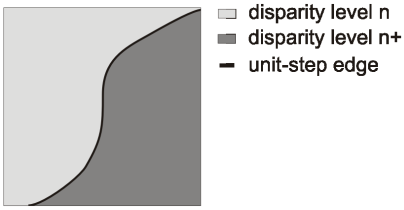


Fig. 7. Detection of unit-step edges.

The proposed technique detects unit-step edges in the disparity map by simple comparison of disparity values in the neighboring pixels. Pixels, whose disparity differs by 1 from neighboring pixels are classified as belonging to an unit-step edge. Those pixels are marked for further processing (Fig. 7). They potentially belong to a false contour in the disparity map.

4.2. Introduction of intermediate disparity levels.

It is supposed that the marked pixels (Fig. 8) should have intermediate values of disparity. So, pixels on both sides of a unit-step edge are processed.

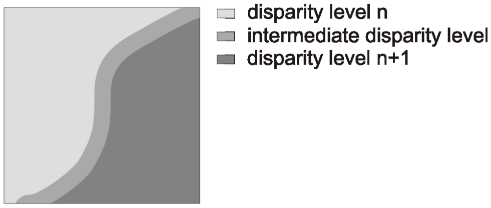
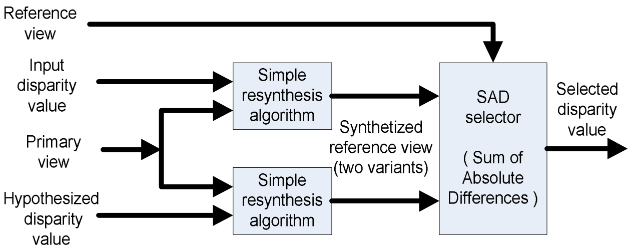


Fig. 8. Intermediate level hypothesis.

4.3. Verification of intermediate level hypothesis.

Unit-step edges may occur in two distinct cases: they may represent actual edges in the scene but they also may result from rough disparity quantization. This decision ambiguity is resolved by verification of hypothesis of intermediate level (Fig. 9). Assumed disparity value is verified by comparison of the quality of the two synthesis variants of the reference view: one obtained from the input disparity values and the other one obtained with the assumed intermediate disparity value. The disparity value that provides better synthesis of the reference view (measured by Sum of Absolute Differences - SAD) is selected as a resultant disparity value.



Reference view
Input disparity value
Primary view
Hypothesized disparity value
Simple resynthesis algorithm
Simple resynthesis algorithm
Synthetized reference view (two variants)
SAD selector
(Sum of Absolute Differences)
Selected disparity value

Fig. 9. Scheme of the verification step.

4.4. Spreading of the hypothesis.

The pixels, that have passed the verification, retain their intermediate disparity values. Then, the mid-level hypothesis is assumed for the neighboring pixels. Thus, the hypothesis is spread to all neighboring pixels within 8-connectivity neighborhoods. These pixels are also marked for further processing (Fig. 10). The mid-level hypothesis is tested for all those pixels.

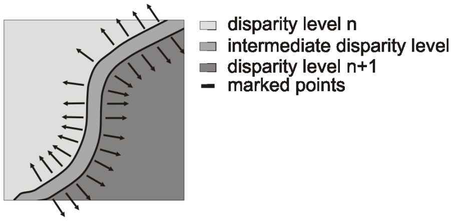


Fig. 10. Spreading direction of intermediate level hypothesis.

4.5. Loop condition.

If there are still marked pixels, algorithm loops to step 2. The algorithm stops when there is no pixel marked for processing. The result of the algorithm is an improved disparity map with new intermediate disparity levels (Fig. 11). Note that usually only a portion of all pixels is processed, i.e. the mid-level hypothesis is verified in the selected pixels only. This observation is closely related to the low complexity of the technique.

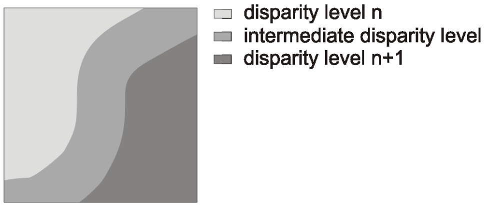


Fig. 11. Refined disparity map.

5. EXPERIMENTAL TESTS

The performance of the proposed technique has been assessed experimentally for both still and moving pictures. In the experiments, input disparity maps with accuracy of the full sampling period, i.e. full-pel maps were refined. These input disparity maps have been obtained with the use of Depth Estimation Reference Software (DERS) [6,7] that employs a state-of-the-art technique based on Graph Cuts optimization algorithm. This software is currently accepted as a reference in MPEG standardization activities.

5.1. Still images

For still images, the proposed technique has been tested using stereo-image datasets published on Middlebury Stereovision page [3]. Percentages of bad pixels have been calculated with respect to the ground-truth depth maps [3,11], and using two values of error threshold: 0.5 and 1.0.

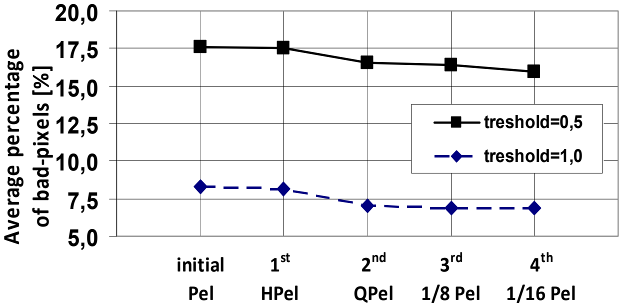
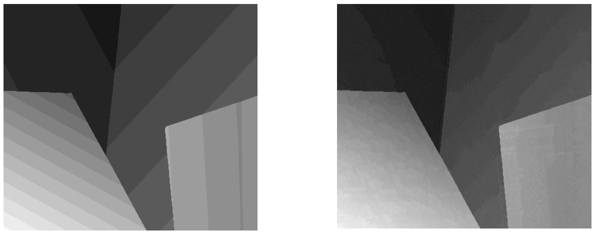


Fig. 12. Improvement of average bad-pixel percentage measure in successive iterations of proposed MLH refinement algorithm for various bad-pixel thresholds: 0.5 and 1.0. The results are averaged over all Middlebury data sets.

The results have been obtained for multiple iterations that correspond to the increasing accuracy of the disparity map up to 1/16th of the sampling period (Fig. 12). For all test images, monotonic decrease of bad-pixel number has been observed, nevertheless improvements in the later iterations are negligible. A depth map improved with the proposed technique is shown in Fig. 13.

We have also compared results of our refinement technique with some state-of-the-art techniques known from

Middlebury website [3]. Tab. 1 presents percentage of bad-pixels averaged over all Middlebury [3] datasets. As the results show, our refinement proposal is competitive to other techniques and may be ranked in the upper middle of the stage. Proposed MLH refinement definitely improves the results of the state-of-the-art DERS (Pel) [6,7] technique with full-pel accuracy, but performs slightly worse than direct application of DERS algorithm with quarter-pel precision – DERS (QPel). This small loss of performance is compensated by lesser computational complexity of MLH technique (see – Section 5.3).



a) before refinement: full-pel precise disparity map

b) after refinement: quarter-pel precise disparity map

Fig. 13. Results of proposed MLH precision refinement algorithm used on Middlebury [3] VENUS dataset.

Table 1. Average percentage of bad-pixels measure [3] of various state-of-the art techniques from Middlebury website [3] compared to MLH (presented mid-level hypothesis algorithm).

Algorithm	Average percent of bad-pixels [%]	
	Threshold = 1.0	Threshold = 0.5
Middlebury - *Adapting BP*	4.23	13.6
Middlebury - *Double BP*	4.19	15.7
Middlebury - *OutlierConf*	4.60	17.3
DERS (QPel)	6.82	15.1
DERS (Pel) + MLH (QPel) **proposed technique**	**7.03**	**16.5**
DERS (Pel)	8.36	17.5
Middlebury - *SAD-IGMCT*	12.5	16.0
Middlebury - *Infection*	20.7	29.4

5.2. Video sequences

As a test material we have used the sequences that are adopted by ISO/IEC MPEG 3DV/FTV expert group [1]. Because no high accuracy ground truth depth map is available yet for those sequences, quality of disparity maps is evaluated indirectly by evaluation of quality of video from a virtual view S (Fig. 14) synthesized in the position in-between of two available views NL and NR. The quality has been measured by comparing the synthetic view S to the real view O taken from the same camera position.

Again, the proposed technique (two iterations) has been applied on the top of the state-of-the-art Depth estimation Reference Software (DERS) [6,7] that works with full-pel accuracy. Luminance PSNR has been calculated for the synthesized view S with reference to the original view O.

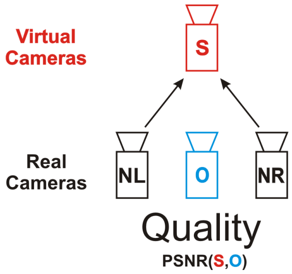


Fig. 14. Setup of experiments for evaluation of performance of the algorithm in case of video sequences.

Table 2. Averaged synthesis quality (for 4 test sequences, mentioned in Section 5.2).

Camera baseline Algorithm	Averaged synthesis quality [dB]		
	default	doubled	both
DERS (QPel)	31.44	32.43	31.67
DERS (Pel) + MLH (QPel) **Proposed technique**	31.91	32.81	32.62
DERS (Pel)	31.67	32.62	32.47

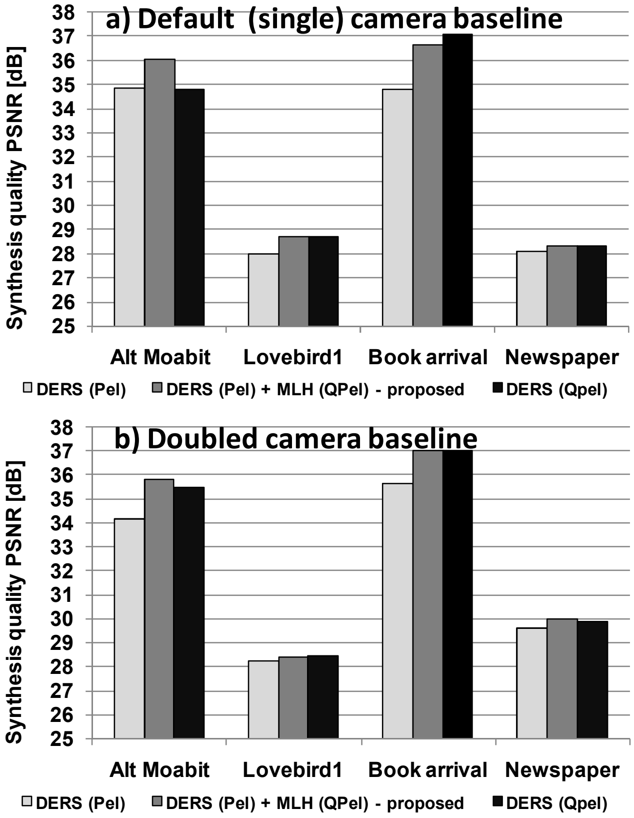


Fig. 15. Comparison of view synthesis quality (PSNR) for tested sequences, various camera baselines.

The results have been compared to those obtained by direct usage of DERS technique with both full-pel and quarter-pel accuracy. The depth estimation in experiments have been performed with various camera baselines: single (default) and doubled.

For 4 test sequences [12-14], the use of the proposed MLH technique (DERS (Pel) + MLH (Qpel)) increases quality of synthesized view from 0.2 to 1.8dB (Fig. 15). For some sequences, proposed refinement technique also outperforms direct DERS (QPel) algorithm – up to about 1dB in case of Alt Moabit sequence (Fig 15a). Although, for some sequences, our approach is outperformed by original quarter-pixel precise DERS (DERS (Qpel)), in average (Tab. 2), the use of proposed technique on the top of the full-pel DERS (DERS (Pel) + MLH (Qpel)) increases quality of synthesized view by about 0.2dB.

5.3. Computational complexity

Computational complexity of the proposed MLH refinement technique was compared against the direct DERS technique (Tab. 3) with use of PC computer (3GHz processor, 4GB of memory). All tests have been performed under the same conditions: processor was not engaged with performing any other background operations etc.

The results show that the direct full-pel estimation and quarter-pel refinement is about 3 times faster than direct quarter-pel estimation using the DERS technique.

Table 3. Average computation time of proposed technique for 4 test sequences, mentioned in Section 5.2.

Algorithm	Depth map computation time [s]
DERS (QPel)	511.6
DERS (Pel) + MLH (QPel) **Proposed technique**	127.4 + **36.3** = 163.7
DERS (Pel)	127.4

6. CONCLUSIONS

We have proposed a simple and robust depth map refinement technique that significantly improves quality of depth maps by providing sub-pixel disparity precision at low computational cost.

For video, the obtained experimental results show that the use of our refinement technique improves the results of the state-of-the-art full-pel precise technique (DERS [6,7]) from 0.2 to 1.8dB of PSNR measured for video sequences. For still pictures, the proposed technique reduces the bad-pixel measure from 1.0 to 1.3 percent points. Considering quarter-pixel precise estimation mode of the reference (DERS) algorithm, our technique provides very competitive quality. The proposed technique does not introduce any artifacts to refined disparity map, because all disparity modifications are done at sub-pixel level. Currently, the performance of the proposed approach is limited by the performance of the available view synthesis techniques.

The main advantage of the proposed approach is that the presented technique is more than 3 times faster than competitive quarter-pixel depth estimation algorithm.

The above described approach is well suitable to 3D video transmission systems where low-accuracy depth maps can be transmitted with low bitrates and they may be refined in the receiver using the proposed fast technique.

7. REFERENCES

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

[2] D. Scharstein, R. Szeliski, „A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International Journal of Computer Vision 2002.

[3] "Middlebury Stereo Vision Page" D. Scharstein, R. Szeliski, http://vision.middlebury.edu/stereo/

[4] A. Klaus, M. Sormann, K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure", International Conference on Pattern Recognition 2006, Hong Kong, China, August 2006.

[5] O. Stankiewicz, K. Wegner, "Depth Map Estimation Software version 3", ISO/IEC MPEG M15540, Hannover, Germany, July 2008.

[6] http://www.tanimoto.nuee.nagoya-u.ac.jp/ - MPEG-FTV web-page, Tanimoto Laboratory, Nagoya University.

[7] M. Tanimoto, T. Fujii, K. Suzuki, "Video Depth Estimation Reference Software (DERS) with Image Segmentation and Block Matching", ISO/IEC MPEG M16092, Lausanne, Switzerland, Feb. 2009.

[8] P. Skulimowski, P. Strumiłło, "Refinement of depth from stereo camera ego-motion parameters", IEEE Electronics Letters, Volume 44, Issue 12, June 5 2008 Page(s):729-730.

[9] A. Banno, K. Ikeuchi, "Disparity map refinement and 3D surface smoothing via directed anisotropic diffusion", The 2009 IEEE International Workshop on 3-D Digital Imaging and Modeling, October 2009, Kyoto, Japan.

[10] T. Pock, C. Zach, H. Bischof, "Mumford-Shah Meets Stereo: Integration of Weak Depth Hypotheses", IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07), Minneapolis, USA, 2007.

[11] D. Scharstein, R. Szeliski, „High-accuracy stereo depth maps using structured light", 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03).

[12] Yo-Sung Ho, E.-K. Lee, C. Lee "Video Test Sequence and Camera Parameters", ISO/IEC MPEG M15419, Archamps, France, April 2008.

[13] I. Feldmann, M. Müller, F. Zilly, et al. „HHI Test Material for 3D Video", ISO/IEC JTC1/SC29/WG11, MPEG M15413, Archamps,France, April 2008.

[14] M. Tanimoto, T. Fujii, N. Fukushima, "1D Parallel Test Sequences for MPEG-FTV", MPEG M15378, Archamps, France, April 2008.