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

ISO/IEC JTC1/SC29/WG11
MPEG/M15540
July 2008, Hannover, Germany

Source	**Poznań University of Technology**
Title	**Depth Map Estimation Software version 3**
Sub group	**Video**

Authors	**Olgierd Stankiewicz** (ostank@multimedia.edu.pl) and
	Krzysztof Wegner (kwegner@multimedia.edu.pl),
	Poznań University of Technology, Chair of Multimedia
	Telecommunications and Microelectronics, Poznań, Poland

1 Introduction

This document presents current state of Poznań University of Technology Depth Map Estimation & View Synthesis Software and is in response to N9468 “Call for Contributions on FTV Test Material” [13], in particular for “Depth Map Estimation & View Synthesis Software” paragraph. Progress of work is described in respect to M15175 [1] and M15338 [2].

2 Overview of changes in the software

Most of the changes since the meeting in Archamps have already been announced on the mail-reflector. Most of them are minor bug-fixes and slight improvements but there is also new improved tool for depth estimation. These changes are described below:

1. **Depth estimation tool memory requirements improvements**
   Memory allocation scheme for belief-propagation based depth-estimation has been revised. New version allocates memory only for currently processed layers, which results in up to 25% lesser memory consumption. Also, message passing structures has been halved with use of smaller data types.

2. **Depth estimation speed improvements**
   Message passing in belief-propagation based depth-estimation employs caching which reduces computational cost by about half.

3. **Minimum disparity option**
   Depth estimation tool has been extended with '-mind' option that selects minimum disparity that is analyzed. Both '-mind' and '-maxd' options allow for arbitrary disparity search range selection. This is useful is cases, where small disparity values are known to be impossible a priori.

4. **Middlebury conformance**
   Previous version of View Synthesis software was accommodated for Middlebury tests, where value 'zero' marks 'unknown' disparity and thus is omitted in synthesis. This prevented synthesis of far background (e.g. in "Dog" sequence). New version fixes that problem.

5. **Computational resolution selection**
   Regardless of performance improvements, belief propagation based depth estimation is still memory consuming process. For slower machines with lesser memory new options have been introduced. '-cw' and '-ch' options allow for user selection of computational resolution. These settings are used during all internal estimation processing. At the end, the results are resampled and rescaled to the original resolution, so the only impact of these options is quality degradation, not the output format or scale.

6. **Bug fixes.**
   Some minor bugs have been fixed.

7. **Dissimilarity metric option for disparity estimation.**
   Additional dissimilarity metric as an alternative for SAD (Sum of Absolute Differences) or SSD (Sum of Squared Differences) metrics. Option '-rank' allows for use of RANK [18], which bring considerable performance gain.

3 Middlebury dataset results

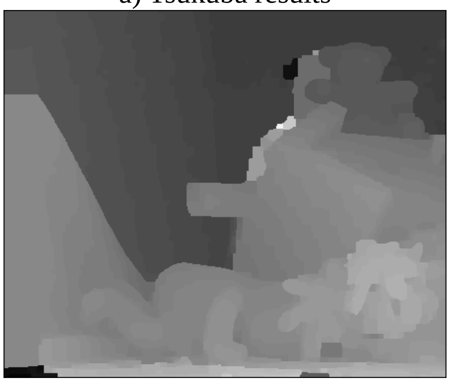
Figure 1 shows disparity maps attained with use of our Disparity Map version 3. As can be seen, there are much less artifacts than in version 2. The only errors that are visible are:

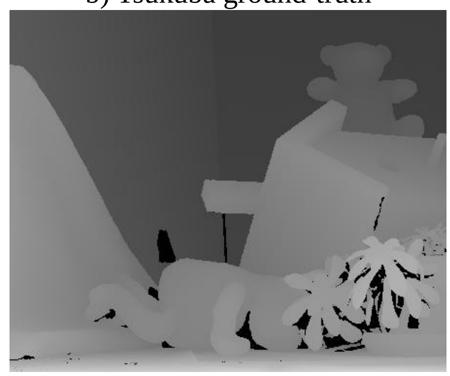
- regions that are occluded or are very nearly to occluded regions (error is propagated across disparity map),
- lack of sub-pixel accuracy (disparity maps are only pixel-accurate).

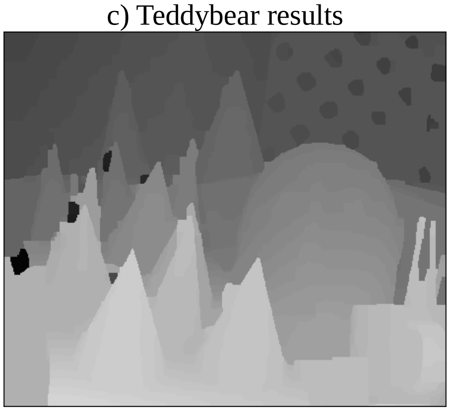
Table 1 shows results by current version of our software in Middlebury tests. Quality of depth map is represented as percentage of pixels with disparity errors (differences between evaluated and ground-truth disparity maps) exceeding given threshold - 'Bad-pixels'. Additionally, we show Normalized Bad Pixel - SAD (*NBP-SAD*) and Normalized Bad Pixel - SSD (*NBP-SSD*) that provide information about magnitude and energy of errors.

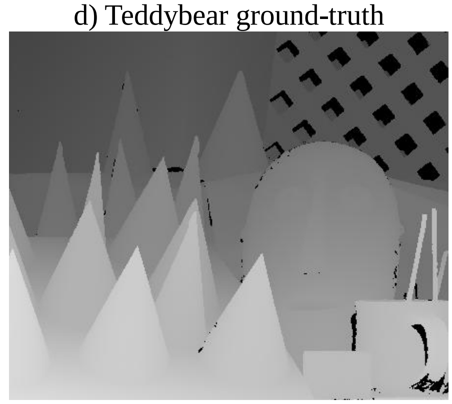

a) Tsukuba results

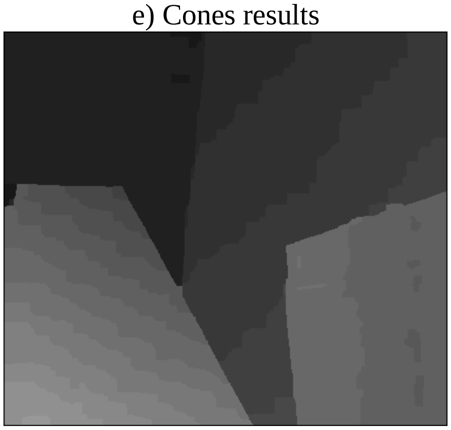

b) Tsukuba ground-truth


c) Teddybear results


d) Teddybear ground-truth


e) Cones results


f) Cones ground-truth


g) Venus results

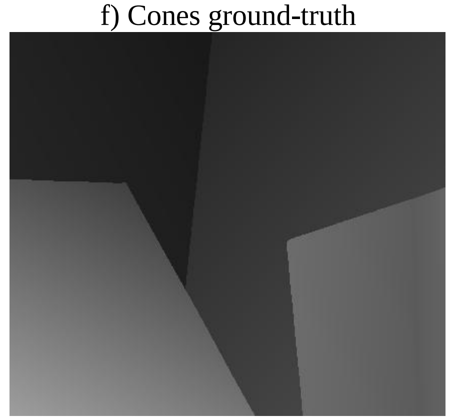

h) Venus ground-truth

Figure 1. *Results of Depth Estimation version 3 (a,c,e,g) and Ground-truth disparity maps (b,d,f,h).*

CONES	Bad Pixels	NBP-SAD	NBP-SSD	PSNR [dB]
Ground truth [15]	0,00%	-	-	31,19
AdaptingBP [4]	2,30%	2,19	5,94	29,01
DoubleBP [5]	3,31%	3,31	15,74	32,51
SubPixDoubleBP [6]	3,20%	3,32	15,97	30,74
AdaptOvrSegBP [7]	3,42%	2,25	6,74	30,14
PlaneFitBP [8]	3,98%	2,77	9,67	31,42
SSD +MF [3]	8,64%	3,12	17,07	30,02
ThreeView [9]	13,54%	2,57	18,94	22,61
PUT-BM+OF (PUT v1) [1]	17,44%	3,61	24,07	28,63
PUT-BP (PUT v2) [2]	5,70%	9,16	234,00	32,93
Our proposal (PUT v3)	**2,81%**	**13,54**	**413,93**	**33,37**

TSUKUBA	Bad Pixels	NBP-SAD	NBP-SSD	PSNR [dB]
Ground truth [15]	0,00%	-	-	35,31
AdaptingBP [4]	1,10%	5,25	35,71	29,62
DoubleBP [5]	0,89%	5,15	33,75	35,32
SubPixDoubleBP [6]	1,22%	4,10	25,46	29,81
AdaptOvrSegBP [7]	1,58%	4,16	24,95	31,51
PlaneFitBP [8]	0,96%	5,38	36,61	34,87
SSD +MF [3]	5,25%	5,28	34,78	33,19
ThreeView [9]	4,65%	3,13	16,38	30,91
PUT-BM+OF (PUT v1) [1]	8,01%	2,50	9,28	29,96
PUT-BP (PUT v2) [2]	1,71%	5,18	49,02	35,19
Our proposal (PUT v3)	**1,52%**	**4,45**	**26,83**	**36,11**

Table 1. *Comparison of current version of software with high-end Middlebury algorithms [10] and algorithms provided to the MPEG.*

NBP-SAD measures value of SAD (evaluated pixel-per-pixel) for 'bad-pixels' and normalized by number of bad-pixels (1).

$$NBP - SAD = \frac{1}{count\, of\ bad\ pixels} \sum_{x,y \in bad\ pixels} \left| G(x,y) - d(x,y) \right| \tag{1}$$

where: *G(x,y)* – ground-truth disparity map *d(x,y)* – evaluated disparity map

NBP-SSD is similar to *NBP-SAD* but measures energy instead of magnitude (2):

$$NBP - SSD = \frac{1}{count\, of\ bad\ pixels} \sum_{x,y \in bad\ pixels} \left| G(x,y) - d(x,y) \right|^2 \tag{2}$$

where: *G(x,y)* – ground-truth disparity map *d(x,y)* – evaluated disparity map

The last column of Table 1 shows results of image resynthesis (right from left) in manner of image distortion measurement with respect to the original – PSNR (Peak Signal-to-Noise Ratio).

As can be seen, proposed algorithm is slightly better that previous version 2 both in Bad-Pixels and PSNR manner. Bad-pixel factor is from 0.2 (for TSKUKUBA) to 3 percent points (CONES) lower than for PUT v2, but still has a distance of about 0.5 percent points to the best performing algorithm. As for PSNR of resynthesis, it still is the best performing algorithm, having even better result of about 1dB.

The score of Disparity Estimation software is about 5 positions higher in Middlebury ranking.

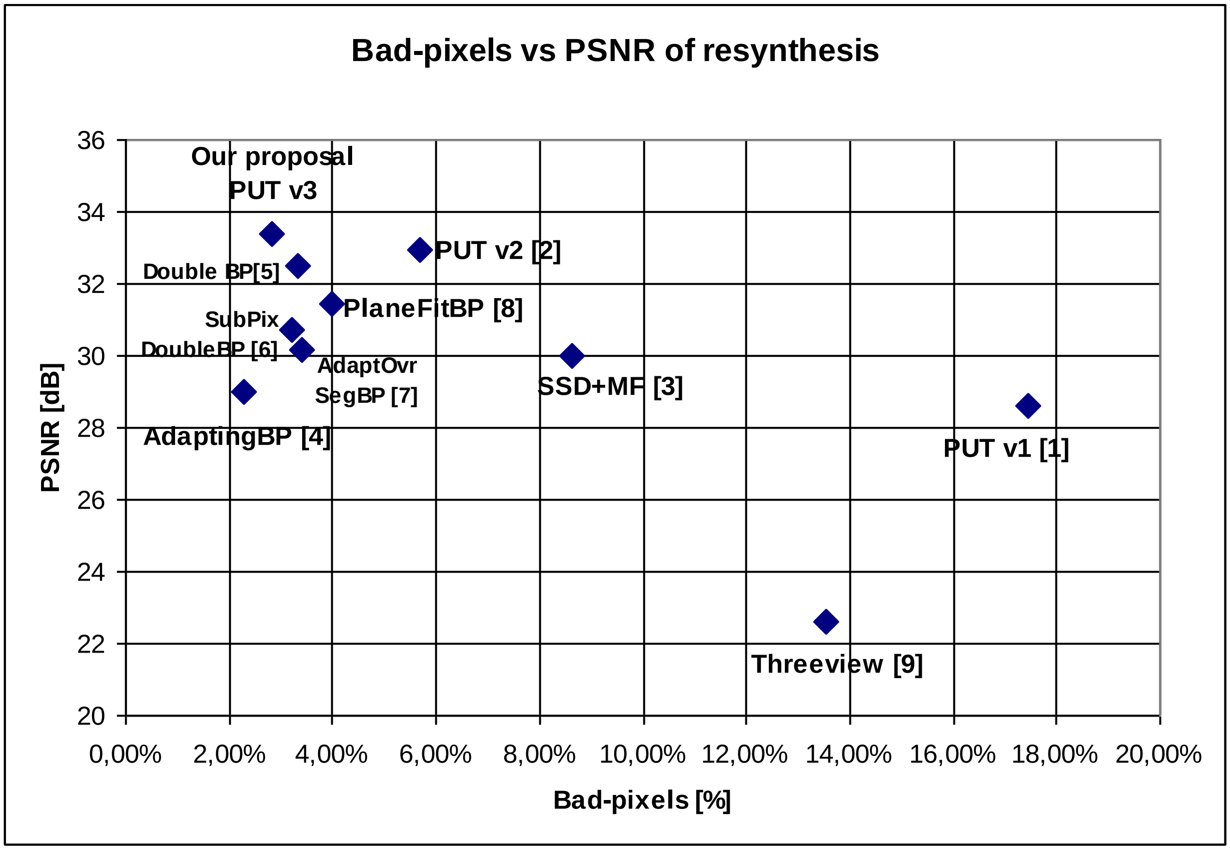


Figure 2. *Ground-truth-based quality (Bad-Pixel) measure versus resynthesis-based (PSNR of resynthesis) on CONES image [10].*

Fig. 2 shows results of competitive algorithms as Bad-pixel versus PSND of resynthesis graph. Good quality of resynthesis (high PSNR) is generally linked with small percentage of bad-pixels, but the relation is ambiguous. For example, methods [4], [6] and [7] are believed to be very efficient regarding to 'bad-pixels' metric [10], but are weak in resynthesis sense. Similarly, [9] is slightly better than [3] according to 'bad-pixels', but is about 3dB worse with respect to PSNR. Taking both PSNR and Bad-pixels into consideration, our proposal lays is left-top corner of the graph, being the best performing algorithm among the competitors.

4 MPEG sequences results

We have performed experiments with some of MPEG sequences [19] to test performance of our software on moving pictures. Unfortunately, the ground-truth disparity maps are unavailable and it was impossible to assess the quality of our algorithm directly. The performance was evaluated with PSNR of view resynthesized in the middle of two adjacent views and corresponding depth maps. Cameras 7 and 9 were chosen as reference views, and sequence for camera 8 was resynthesized and compared with the original.

a) original sequence b) estimated disparity map

Figure 3. “Leaving Laptop” [19] sequence.

a) original sequence b) estimated disparity map

Figure 4. “Alt Moabit” [19] sequence.

Figures 3 and 4 show results of disparity estimation. There are less artifacts than before, for example the edges of objects are estimated precisely.

Sequence	**PSNR of resynthesis (v2) [dB]**	**PSNR of resynthesis (v3) [dB]**
Leaving Laptop	28,27	28,76
Flowers	28,01	28,82
Alt Moabit	30,85	32,53

Table 2. *PSNR of resynthesized view - comparison of current version of software with previous version on exemplary MPEG sequences.*

Table 2 presents comparison of Poznan software version 2 with version 3. As can be seen, the PSNR of resynthesized view (camera 8) is slightly better (for about 1dB) than previously.

Unfortunarelly, still there are some problems that are yet to solve. E.g. disparity of floor of the room in “Leaving Laptop” sequence (Figure3 b) is estimated wrongly. One can notice that depending on the noise it dithers in time between the correct and wrong value.

5 Conclusions

In this paper we have presented a new version (version 3) of experimental Depth Estimation & View Synthesis Software. The improvements have been briefly described and results have been discussed.

In particular, new version of Depth Estimation software is about twice as fast as version 2, it has lesser memory requirements and provides disparity of higher quality.

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”, MPEG 2008/M15338, Archans, Framce, April 2008.

[3] D. Scharstein, R. Szeliski. „A taxonomy and evaluation of dense two-frame stereo correspondence algorithms”, IJCV 2002.

[4] A. Klaus, M. Sormann, K. Karner, „Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure”, International Conf. on Pattern Recognition 2006.

[5] Q. Yáng, L. Wang, R. Yang, et al., „Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling”. Computer Vision and Pattern Recognition 2006.

[6] Q. Yang, R. Yang, J. Davis, D. Nistér, “Spatial-depth super resolution for range images”. Computer Vision and Pattern Recognition 2007.

[7] Anonymous, “Stereo reconstruction with mixed pixels using adaptive over-segmentation”, Computer Vision and Pattern Recognition 2008.

[8] Anonymous, “Near real-time stereo for weakly-textured scenes”. Computer Vision and Pattern Recognition 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] “Middlebury Stereo Vision Page, http://vision.middlebury.edu/stereo/

[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

[15] D. Scharstein and R. Szeliski, „High-accuracy stereo depth maps using structured light” Computer Vision and Pattern Recognition 2003, volume 1, pages 195-202, June 2003.

[16] “Middlebury Stereo Vision Page, http://vision.middlebury.edu/stereo/

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

[18] R. Zabih, J. Woodfill, “Non-parametric local transforms for computing visual correspondence”, Proceedings of the third European conference on Computer Vision (Vol. II) , Stockholm, Sweden, pages: 151 – 158, 1994.

[19] Feldmann, M. Mueller, F. Zilly, R. Tanger, K. Mueller, A. Smolic, P. Kauff, T. Wiegand **„**HHI Test Material for 3D Video”, MPEG 2008/M15413, Archans, Framce, April 2008.