**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/M15338**
**April 2008, Archamps, France**

**Title** **Depth Map Estimation Software version 2**
**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 our Depth Map Estimation 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].

# 2 New tools in the software

We present achievements that have been made in our depth estimation software since [1]. This paper contains some thoughts, some new tools and some proposals for performance analysis. In particular we introduce:

- paradigm of depth map quality measurement by view resynthesis,
- view synthesis tool,
- depth map quality measures exploiting ground-truth map,
- depth map quality measurement by view resynthesis,
- Belief Propagation based depth estimation tool,
- novel accuracy refinement algorithm.

All included applications have 'help screens' which are displayed in console window after running. In case of any questions please contact authors via e-mail.

# 3 View synthesis

Works on depth map estimation, in context of multiview video relay, should be conducted with view synthesis kept in mind. The end-user of the multiview system never sees depth map, so we conclude that quality of depth maps should be measured with respect to the quality of synthesis, instead of quality of depth map itself. Quality measurement methods constructed on synthesis/resynthesis paradigm are also useful in evaluation of goal function for wide spectrum of optimization algorithms. Therefore, we focused on view synthesis method that would:

- provide good quality of the synthesized views,
- allow depth map quality analysis,
- provide feedback mechanism for optimization algorithms.

We propose a very simple synthesis scheme that can be used only for stereo pairs which come from linearly positioned camera array. Scope of proposed tool is to assess quality of depth map and also to provide some reference for other more complex synthesis algorithms.

Our algorithm exploits two disparity maps and corresponding reference views and synthesizes view in between of them. Pixels from reference views are translated (with respect to their disparity value) and are weighted together (Fig. 1). Weights depend on position of synthesized view in between reference views. Because both of the views are used during the synthesis, most of occluded regions in the synthesized scene are covered.

Proposed algorithm uses disparity maps instead of depth maps, which reduces complexity of the task in case of stereo pairs.

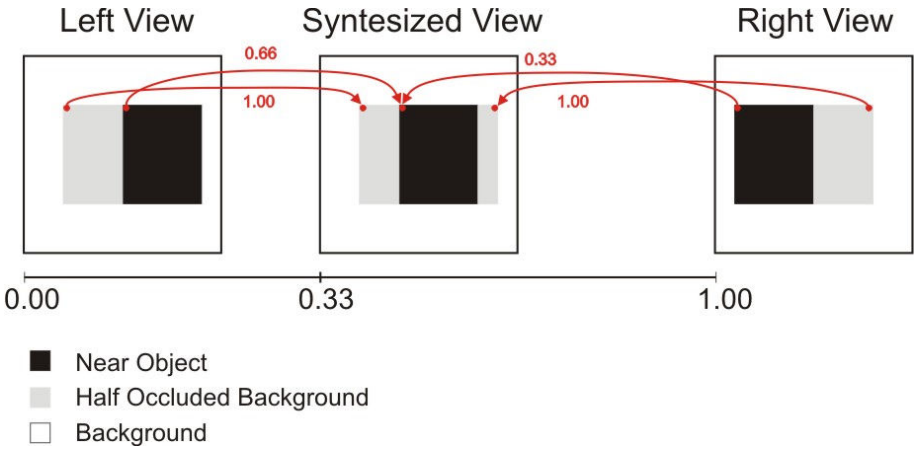



Figure 1. *Synthesis of an intermediate view in between of left and right view (⅓ from the left).*

## 4 Depth map quality measurement

One of the most common methods for measurement of quality of depth maps is 'bad-pixels' metric [10]. Quality of depth map is represented as percentage of pixels, whose disparity errors (differences between evaluated and ground-truth disparity maps) exceeds given threshold. 'Bad-pixels' metric is simple and straight-forward, but it is not sufficient, because it does not convey information about magnitude nor energy of the errors. We propose two supplementary metrics: Normalized Bad Pixel - SAD (*NBP-SAD*) and Normalized Bad Pixel - SSD (*NBP-SSD*) that provide missing information.

*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):

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

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

Still, quality measurement methods based only on comparison of estimated depth map and ground-truth map are flawed, because:

- ground-truth maps are not available for video-sequences, so these methods are not applicable,
- depth maps are ambiguous in occurrence of specular highlights, reflections or transparent objects,
- aim is not to provide a good disparity map but to have an ability of providing good view resynthesis.

We propose that measurement of depth map quality should be done with respect to resynthesized view quality: in particular, with use of the most common method of objective image distortion measurement - PSNR.

Fig. 2 shows unconformity of ground-truth-based quality (Bad-Pixel) measures with resynthesis-based (PSNR of resynthesis). 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.

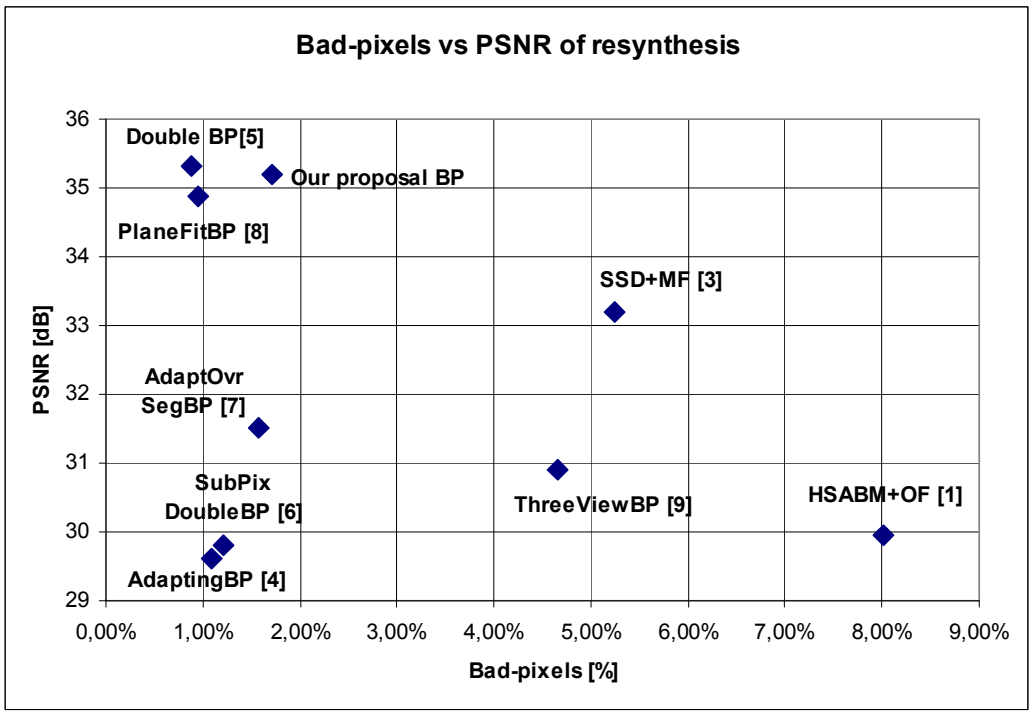



Figure 2. *Ground-truth-based quality (Bad-Pixel) measure versus resynthesis-based (PSNR of resynthesis) on Tsukuba Data Set [10].*

# 5 Detailed description of depth map estimation tools

In [1] we have proposed hybrid approach for the depth map estimation problem. Solution, that has been presented there, exploits modified optical-flow algorithm as the main iterative computation core and hierarchical shape-adaptive block matching for the first guess of disparity map.

This paper extends functionality of previously presented software by allowing use of alternative first guess estimation (Belief Propagation based versus previous block matching based) and by presenting tool for disparity refinement (by mid-level hypothesis).

## 5.1 Belief propagation based estimation

Belief propagation (BP) is an optimization algorithm widely used in computer vision tasks. It can be successfully used in context of depth estimation [14]. Disparity map is modeled as 2-dimensional Markov-field. Nodes on the mesh communicate with others by message passing mechanism (Fig. 3).

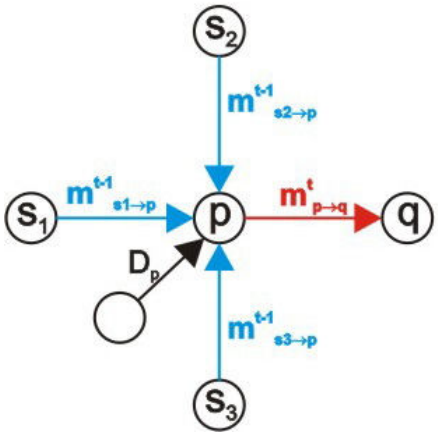



Figure 3. *Message passing in BP algorithm, where:* $m^t_{s\to d}$ – message passed in t-th iteration from node *s* to node *d*, $V(f_p,f_q)$ – cost of belief change from disparity $f_p$ to disparity $f_q$.

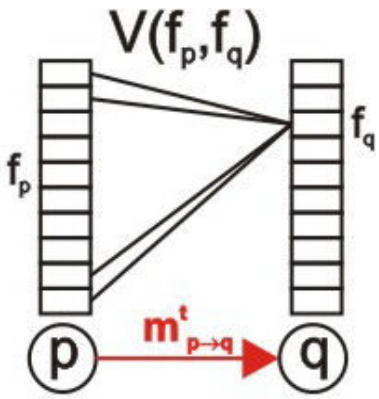



Figure 4. *Transfer of single message containing information about all possible disparities.*

Messages contain information about beliefs of each node, specifically – beliefs about disparities (Fig. 4). Beliefs attained from neighboring nodes and self beliefs are mixed together to produce new beliefs of the node. The process of message passing is repeated until convergence. In the end, the beliefs with the highest likehood are chosen as a final result.

Some specific details about the devised algorithm are below:

- hierarchical processing in layers (disparity maps are estimated from the lowest resolution to the full resolution in coarse-to-fine manner),
- 4 neighbors are considered in message passing in single layer (left, right, up, down),
- uses pot, linear and quadratic model of belief passing cost function of BP algorithm,
- pixel differences (1-point SAD) used as observations.

Notice:

- BP computation cost is highly dependent on resolution and accuracy of the disparity map,
- Computation cost related with accuracy can be lowered at BP stage; accuracy can be improved later by other inexpensive methods,
- hierarchical approach reduces number of iterations.

### 5.2 Disparity map refinement by mid-level hypothesis

The aim of refinement algorithm is to improve accuracy of input disparity map. The number of levels is increased by addition of intermediate levels between existing ones. Hypothesis of existence of intermediate level is put over unit-step edges in input disparity map (Fig. 5).

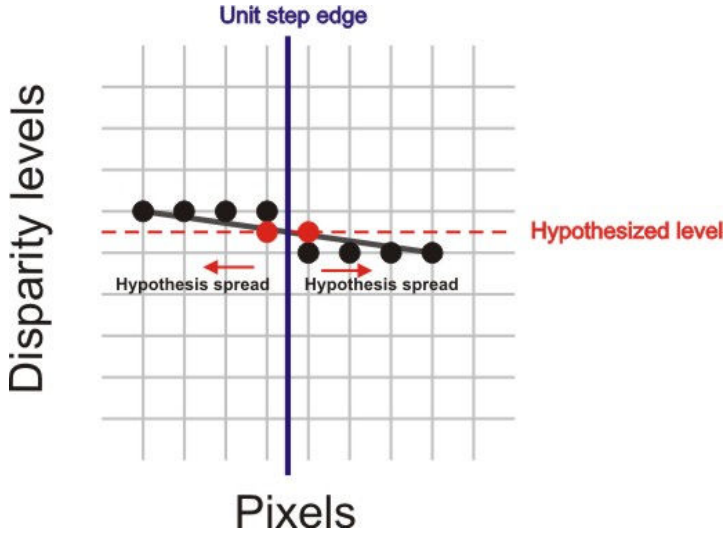



Figure 5. *Spreading of level hypothesis starting from unit-step edge.*

*Unit-step edges exist over flat regions slightly sloped to the camera with roughly estimated depth. They are considered to be important because they introduce synthesis artifacts (Fig. 6 a,b) in spite of other surfaces: flat regions in parallel to the camera are not vulnerable to low accuracy; surfaces highly angled suffer from lack of spatial resolution and not of lack of accuracy (Fig. 6 c,d).*

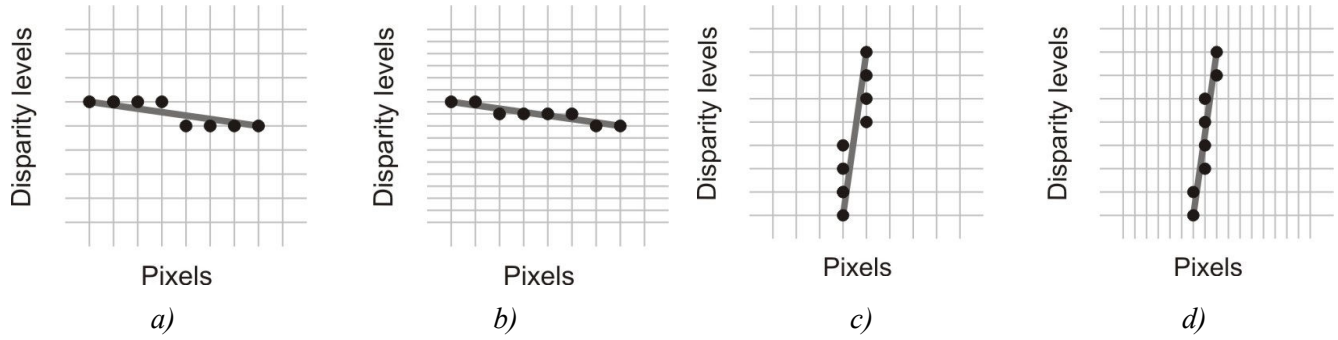



Figure 6. *Different angles of surfaces and relevant direction of resolution improvement.*
*a,b – surface nearly parallel to the camera* *c,d – surface perpendicular to the camera*
*a,c – without improvement* *b,d – with improvement in relevant direction*

The hypothesis is verified by comparison of resynthesized view (attained with use of the processed disparity map) with original view. Views are resynthesized with scheme presented in point 3. Resynthesized views are compared with respect to SAD (Sum of Absolute Differences) metric.

One iteration of the proposed algorithm is composed of the following steps:

1. Single-level disparity edges are found.
2. Intermediate level is hypothesized along these edges.
3. The hypothesis is verified with respect to SAD metric of resynthesized view.
4. Points that have passed verification spread hypothesis to their neighbors.
5. Process continues as long as any point passes the verification.

Single use of the algorithm doubles the number of disparity levels. Successive use can lead to even greater improvement by factor of ×4 - ×8. Figure 7 shows fall of bad-pixel count, relative to input disparity map (×1) in percents. Improvement progresses to about 82% of initial bad-pixel count at third iteration (×8) and is stopped (at ×16) by artificial rugged edges introduced by noisy verification scheme.

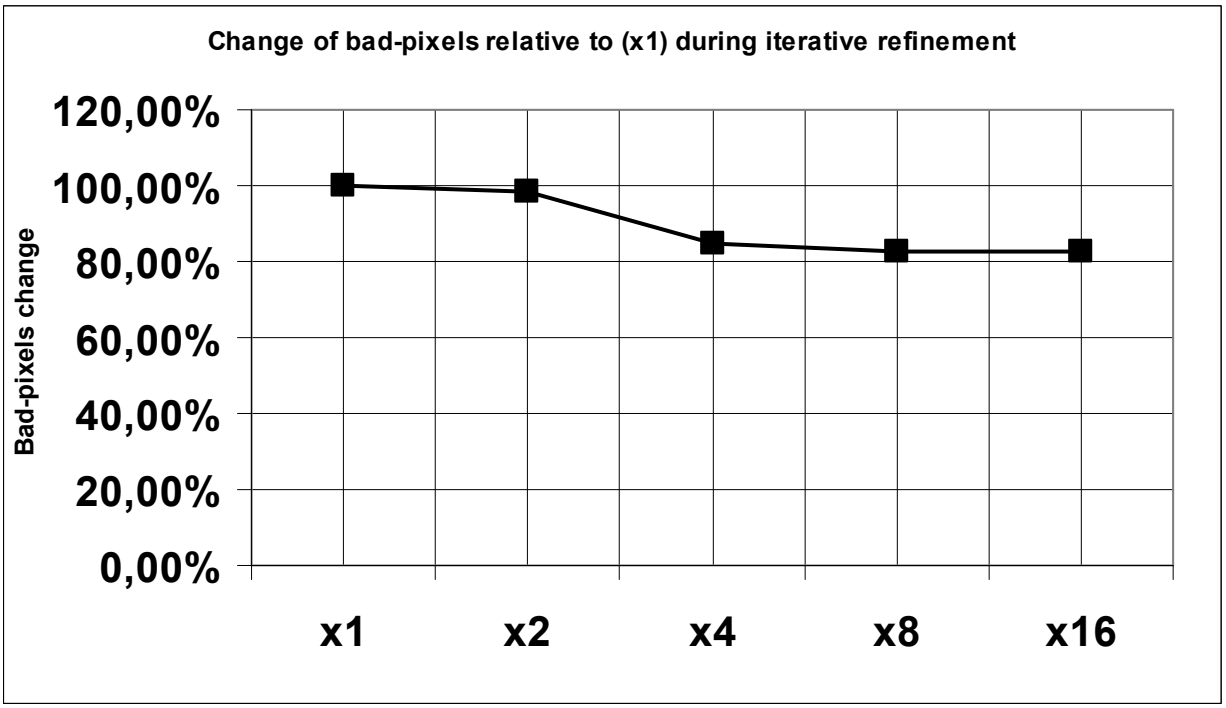



Figure 7. *Change of bad-pixel, relative to initial disparity map (×1 - 100% of bad-pixels) in percents, during iterative refinement.*

# 6 Results

The results of described tools and performance of proposed quality measures are presented below.

| CONES | Bad Pixels | NBP-SAD | NBP-SSD | PSNR [dB] |
|---|---|---|---|---|
| **Ground truth [2]** | 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 |
| **HSABM+OF [1]** | 17,44% | 3,61 | 24,07 | 28,63 |
| **Our proposal BP** | **5,70%** | **9,16** | **234,00** | **32,93** |

| TSUKUBA | Bad Pixels | NBP-SAD | NBP-SSD | PSNR [dB] |
|---|---|---|---|---|
| **Ground truth [2]** | 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 |
| **HSABM+OF [1]** | 8,01% | 2,50 | 9,28 | 29,96 |
| **Our proposal BP** | **1,71%** | **5,18** | **49,02** | **35,19** |

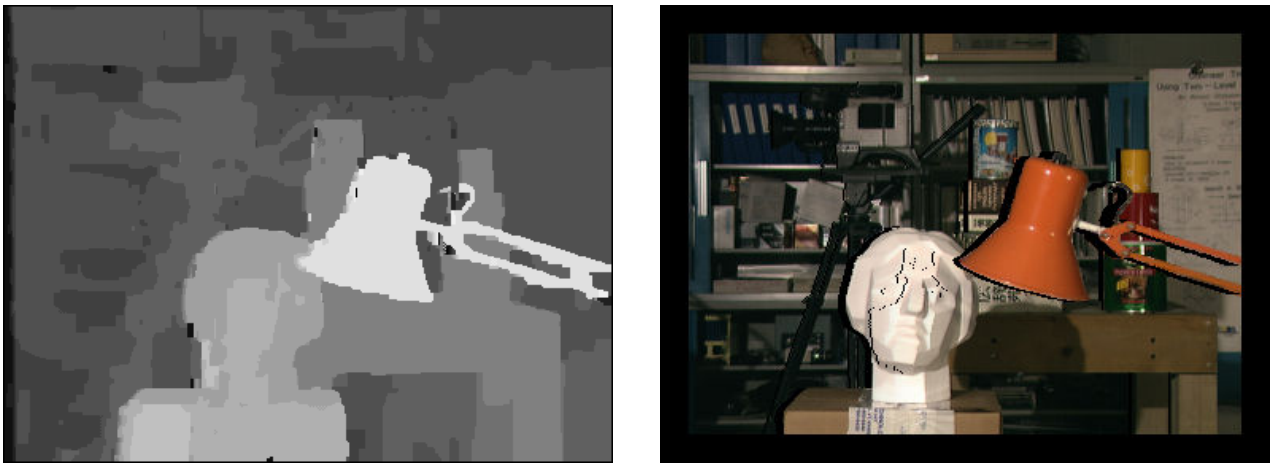
Table 1. Comparison of proposed algorithm performance with high-end Middlebury algorithms [10] and algorithms provided to the MPEG.

Table 1 shows comparison of performance (expressed with respect to proposed quality measures) of selected high-end algorithms, whose results are available at Middlebury [10], with algorithms provided to MPEG and with our proposed algorithm. According to the bad-pixels measure, our proposal is in upper-middle of the score, having distance of about 1-2 percent points to the best one. According to PSNR measure, it is the best for the CONES dataset, and has distance of about 0,1dB to the best one for the TSUKUBA dataset.

a) b)

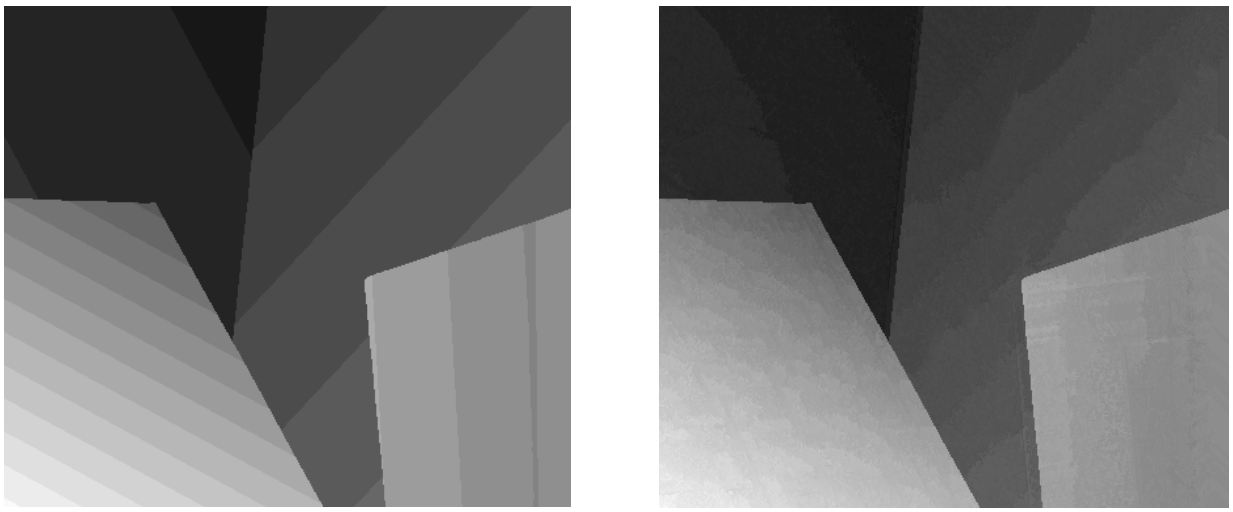
Figure 8. *Results of described synthesis algorithm*
*(a) input disparity map b) view synthesized in between two reference views.*

Figure 9 presents results of our complete depth processing tool chain. The depth has been estimated for TSUKUBA dataset. Some artifacts (for example at power cord of a lamp) result from hierarchical nature of devised BP algorithm. Small details has been that omitted by estimation process.



*a)* *b)*

Figure 9. *Results of described BP depth estimation algorithm (a) and results of synthesis algorithm (b) with occlusion mask (black).*



*a) before refinement (*×1*): per pixel disparity accuracy*

*b) after refinement (*×8*):* ⅛ *sub-pixel disparity accuracy*

Figure 10. *Results of proposed accuracy refinement algorithm on VENUS dataset.*

Results of presented accuracy refinement tool (Fig. 10) show that proposed mid-level hypothesis algorithm performs very well. Disparity map over all surfaces has been significantly improved. Note that devised algorithm introduced no new artifacts, even at untextured regions (upper-right corner).

## 7 Conclusions

In this paper we have presented a new version (version 2) of experimental Depth Estimation Software. The improvements has been described both from technical and conceptual point of view.

In particular, we considered quality measurement problem with respect to multi-view applications, which led us to introduction of view synthesis tool. View synthesis algorithm is straight-forward and simple, because its goal is mainly to assess quality of depth maps, both for performance measurement and for internal algorithm operation.

Also, belief propagation based tool for depth estimation has been introduced. It offers considerably better performance than previous one (hierarchical, shape-adaptive block matching). Appropriate results, considering quality measurement, have been attached.

Finally, new technique for disparity map refinement has been proposed. *Refinement by mid-level hypothesis* turned out to be robust and computationally efficient method for providing significant improvement in disparity map accuracy.

The executable versions of all devised tools have been attached to the document package. The source code of the software will be made available on demand via e-mail.

## 8 References


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

[2] 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.

[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, 25(7):787–800, 2003.