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/M21365
July 2010, Torino, Italy

Title	**Improved depth estimation with advanced occlusion handling**
Sub group	**Video**
Authors	**Krzysztof Wegner** (kwegner@multimedia.edu.pl), **Olgierd Stankiewicz** (ostank@multimedia.edu.pl) Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland

1 Introduction

Good and reliable depth estimation is the key for the success of 3DV/FTV standardization. This contribution reports an experimental results of automatic depth estimation based on novel occlusion handling technique in DERS. This method was used for estimated currently best available depth maps for Poznan sequences [2].

2 Depth estimation with occlusion handling

The main problem with depth estimation for Poznan sequence [2] are large disparity gaps in the scene which cause occlusion in neighboring views.

Current version of DERS consists basically of two steps: matching cost computation and graph-cut optimization. Firstly for each pixel in the center view, matching cost of given pixel to all of the pixels in the left- and right views along the epipolar line is calculated. Then the most likely disparity value for every pixel is obtained by minimization of a cost energy function by graph-cut algorithm [3] The energy function consists of a similarity term (the matching cost), and a smoothing term.

Nevertheless due to occlusion some of pixels aren’t visible in some (may by in all) of the reference views. To handle occlusions in current version of DERS, the similarity term (matching cost) is calculated by matching between the center and left views, and the center and right views, and then the smallest cost is simply selected. This strategy is called min-selection. In case of sequences with low texture and high disparity range, like i.e. Poznan sequences, this strategy may case errors in depth maps. Those errors mainly occur near spatial edges of objects in the foreground. Such errors are mainly caused by selected similarity term (matching cost) of pixels from wrong view (analyzed pixels are in fact occluded in chosen view). Such pixels have lesser noise or they are more similar to analyzed pixels then correct ones.

There were some tries to overcome this problems. For example by color-independent occlusion detection [4]. This algorithm was successive used for improved depth maps for some of 3DV test sequences in the past especially for Poznan Hall1 and Poznan Hall2. Despite of the name of the algorithm it is based on color similarity in the views to estimate occluded pixels.

In this contribution we propose an new technique to iteratively estimate occlusions in reference views along with depth estimation, and use of estimated occlusion map for similarity term calculations.

To handle occlusions in our proposal the similarity term (matching cost) is calculated based on matching between the center and left views, and the center and right views, and then the non-occluded one is selected. (current DERS use min-selection strategy)

To estimate occlusion we projected(synthesized) estimated depth map onto neighbor views and then test each pixels new disparity candidate against it to indicate whether it is occluded or not.

Implementation

The similarity term (matching cost) in our algorithm is calculated and buffered independently for each of reference views. In comparison current version of DERS calculated and buffered min-selected matching cost, it means that for given pixel and disparity tested, best matching cost among all reference views is selected.

Next in iterative procedure of graph-cut at each step when considering given particular disparity, matching cost including occlusion is selected based on previously calculated matching cost to all of the reference views and currently estimated occlusion in each reference views. Matching cost is selected as best non-occluded matching cost for given disparity from all reference views.
For occlusion/non-occlusion test, we project currently considered disparity at given point onto each reference views and compare with currently stored depth for that reference view. If projected disparity is further then currently stored value for that view, tested point is considered as occluded in that tested reference view.

Occlusion estimation is done after each iteration of graph cut algorithm. Currently estimated depth is projected (synthesized) onto each reference views, z-tested, and stored as depth map of that reference view.

Rest of the depth estimation algorithm remains unchanged.

Complicity and Memory analysis

Current version new algorithm requires doubled amount of memory for buffering of the matching cost, because it is stored for both reference views (in current version, it was stored once - only one min-selected matching cost was buffered).

Current version of our algorithm also estimated depth slower that the original DERS, due to matching cost selection based on occlusion, which is done in every iteration of the graph cut algorithm. Another slowing factor is occlusion estimation.

3 Results

We have compared our proposal against current (original and not modified) version of DERS. We use standard procedure (fig. 1) that has been used in numbers of exploration experiment [1] so far. We have estimated depth maps for two views and then we have synthesized view in between and compared (in terms of PSNR) that synthesized view with the original view (captured by a real camera). Detailed views number are given in table 1.

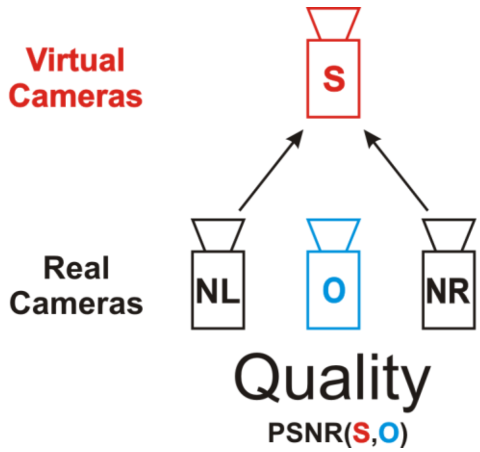

Figure 1. Test setup.

Table 1. The specification of views numbers for experiment.

Data set	**Original Pair NL-NR**	**Synthesized View S (O)**
Poznan_Hall1	3-1	2
Poznan_CarPark	5-3	4
Book Arrival	9-7	8

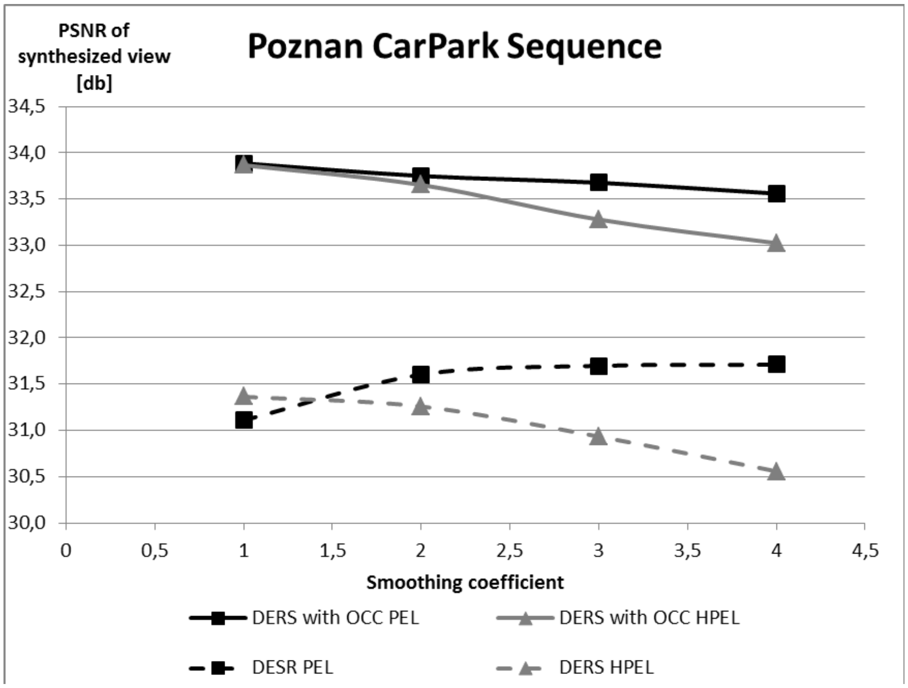

Figure 2. Performance comparison of reference DERS and our proposal - modified DERS in terms of quality of synthesized view against smoothing coefficient.
Poznan CarPark sequence.

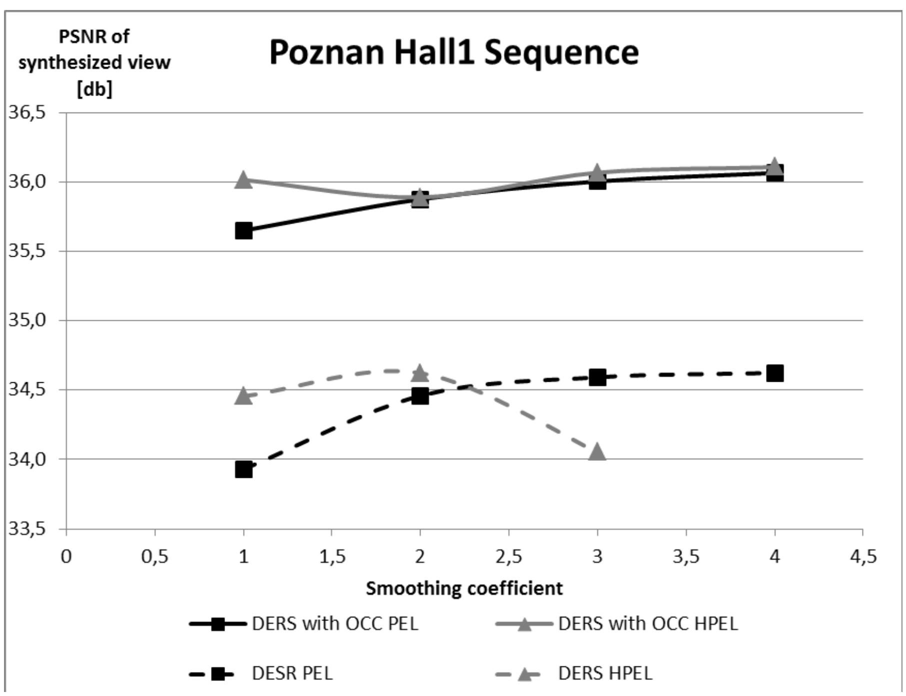

Figure 3. Performance comparison of reference DERS and our proposal - modified DERS in terms of quality of synthesized view against smoothing coefficient.
Poznan Hall1 sequence

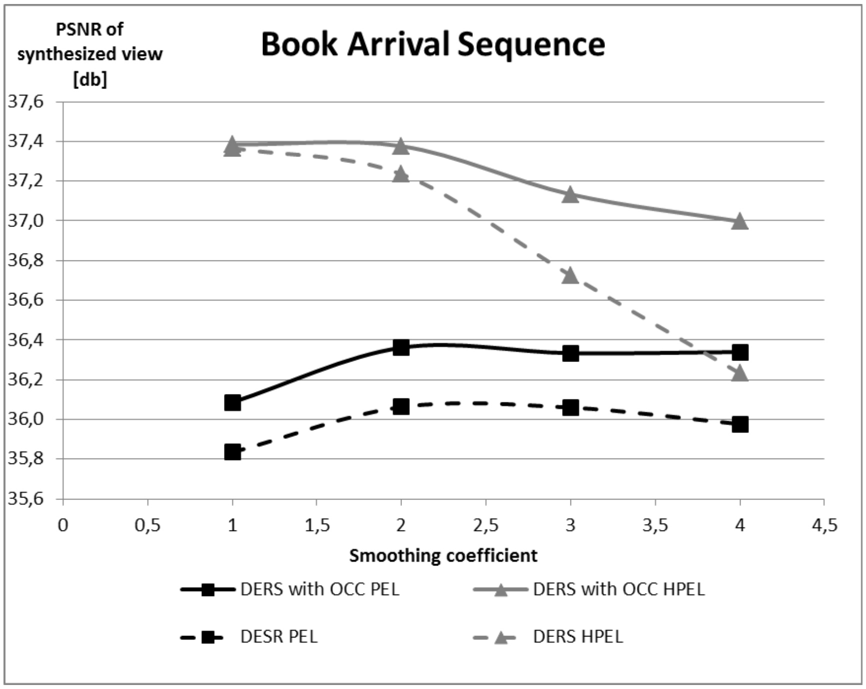

Figure 4. Performance comparison of reference DERS and our proposal - modified DERS in terms of quality of synthesized view against smoothing coefficient.
Book Arrival sequence

4 Conclusions

New technique for automatic depth estimation can be used in various DERS configuration.
As our experiments revealed, that our new technique improves quality of depth maps up to 2,5 dB (in terms of PSNR of synthesized view).

We recommend to include our new technique into DERS.

5 References

[1] “Description of Exploration Experiments in 3D Video Coding” ISO/IEC JTC1/SC29/WG11 MPEG 2009/W11630, Guangzhou, China, October 2010.
[2] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznań Multiview Video Test Sequences and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, China, October 2009.
[3] “Report on Experimental Framework for 3D Video Coding”, ISO/IEC JTC1/SC29/WG11, MPEG 20010 / N11631, Guangzhou, China, October 2010.
[4] Takanori Senoh, Kenji Yamamoto, Ryutaro Oi, Yasuyuki Ichihashi, Taiichiro Kurita, “Depth Estimation Experiment with Poznan Hall”, ISO/IEC JTC1/SC29/WG11 MPEG2010/M18221 October 2010, Guangzhou, China