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/M16027
February 2009, Lausanne, Switzerland

Title	**Analysis of sub-pixel precision in Depth Estimation Reference Software and View Synthesis Reference Software**
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 results of experiments performed with depth estimation and view synthesis software, kindly provided by Nagoya University [1]. The current version of MPEG FTV 3DTV reference package was used. The results of experiments were used for analysis of sub-pixel precision mode and its influence on performance of Reference Software.

As shown, despite the findings of the group in "AHG on 3D Video and FTV Coding" [2], the sub-pixel precision mode of depth-estimation reference software has almost no impact on quality of synthesized view. The gain observed during the 3DTV Exploration Experiments comes from sub-pixel precision mode in synthesis software. Therefore, there is no reason for using currently implemented sub-pixel precision mode in depth estimation process.

2 Experiments setup

All the experiments were performed according to guidelines for Exploration Experiments in "Description of Exploration Experiments in 3D Video Coding" [3]. Quality of depth estimation and view synthesis software setup was evaluated by quality of two views (SL, SR), synthesized from side-views (NL, NR) compared to quality of original views (OL, OR) (Figure 1). The steps of each experiment were as follows:

1. **Estimate depth maps** for two side-views NL and NR from neighboring views (for example NL-1, NL, NL+1 and NR-1, NR, NR+1 respectively, for camera distance equal to 1).
2. **Synthesize views** SL and SR placed at positions of OL and OR with use of NL+depth and NR+depth.
3. **Compare synthesized views** SL,SR with original views OL,OR subjectively and by PSNR.

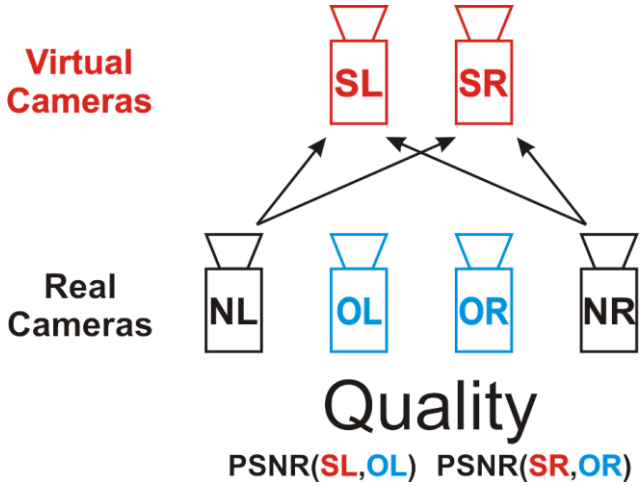

Figure 1. Setup of experiments for depth-estimation/view-synthesis software evaluation.

Due to limitations of computational power, only a few of MPEG 3DTV test sequences were chosen for experiments:

- 'Outdoor Alt Moabit' sequence (kindly provided by HHI),
- 'Book Arrival' sequence (kindly provided by HHI),
- 'Newspaper' sequence (kindly provided by GIST),
- 'Dog' sequence (kindly provided by Nagoya University),
- 'Lovebird 1' sequence (kindly provided by ETRI).

It proposed to perform similar tests on remaining sequences.

For each of the sequences:

- DERS was used to produce depth maps with various configuration parameters values (camera distance, smoothing coefficient) in Pel, Hpel and Qpel estimation precision modes.
- VSRS was used to generate synthesized views from all produced depth maps with use of various synthesis precision modes: Pel, HPel and Qpel.
- The results were averaged over all used Depth Estimation Reference Software configuration parameter values.
- The results were presented as a function of depth estimation and view synthesis software precision modes.

3 Results

Table 1. Averaged PSNR of synthesized virtual views (left and right) for 'Newspaper' sequence for various pixel precisions.

Performance [PSNR]		Depth Estimation Precision		
		Pel	Hpel	Qpel
View Synthesis Precision	Pel	29,6761	-	-
	Hpel	30,0276	29,8861	-
	Qpel	30,0023	29,8539	28,2948

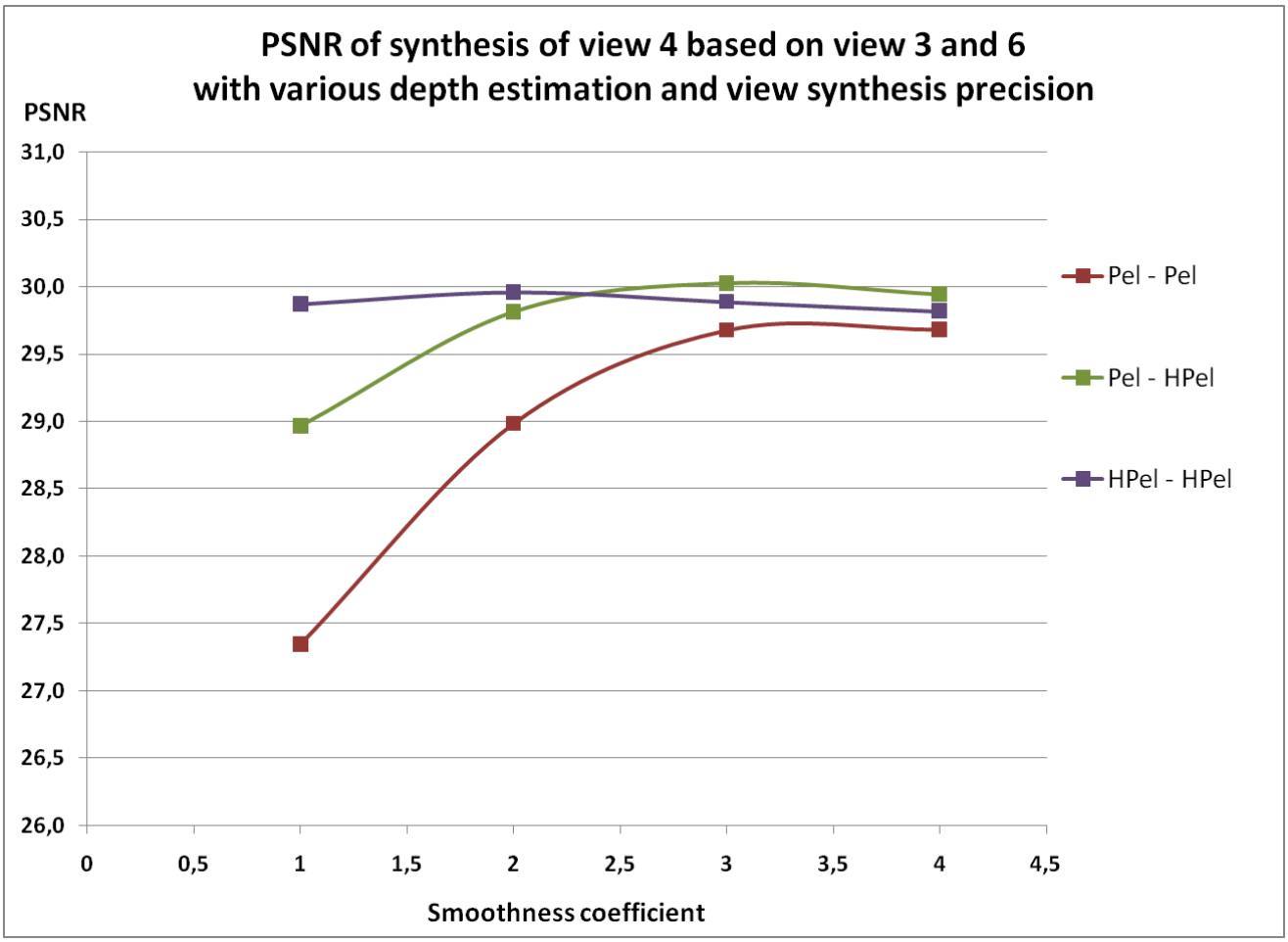

Figure 1. PSNR of synthesized virtual views (left and right) for 'Newspaper' sequence for various precision modes (depth estimation precision – view synthesis prediction).

As can be noticed in Table 1 and Figure 1, precision of depth estimation (Pel-HPel versus HPel-HPel) makes difference only in case of small values of smoothing coefficient, which are not recommended for Newspaper sequence. In range of recommended smoothing coefficient, where PSNR reaches its plateau, there is little evidence that precision of depth estimation gives any gain.

Table 2. Averaged PSNR of synthesized virtual views (left and right) for 'Book arrival' sequence for various pixel precisions.

Performance [PSNR]		Depth Estimation Precision		
		Pel	Hpel	Qpel
View Synthesis Precision	Pel	35,659	-	-
	Hpel	36,9301	36,2315	-
	Qpel	36,4471	36,1465	32,2936

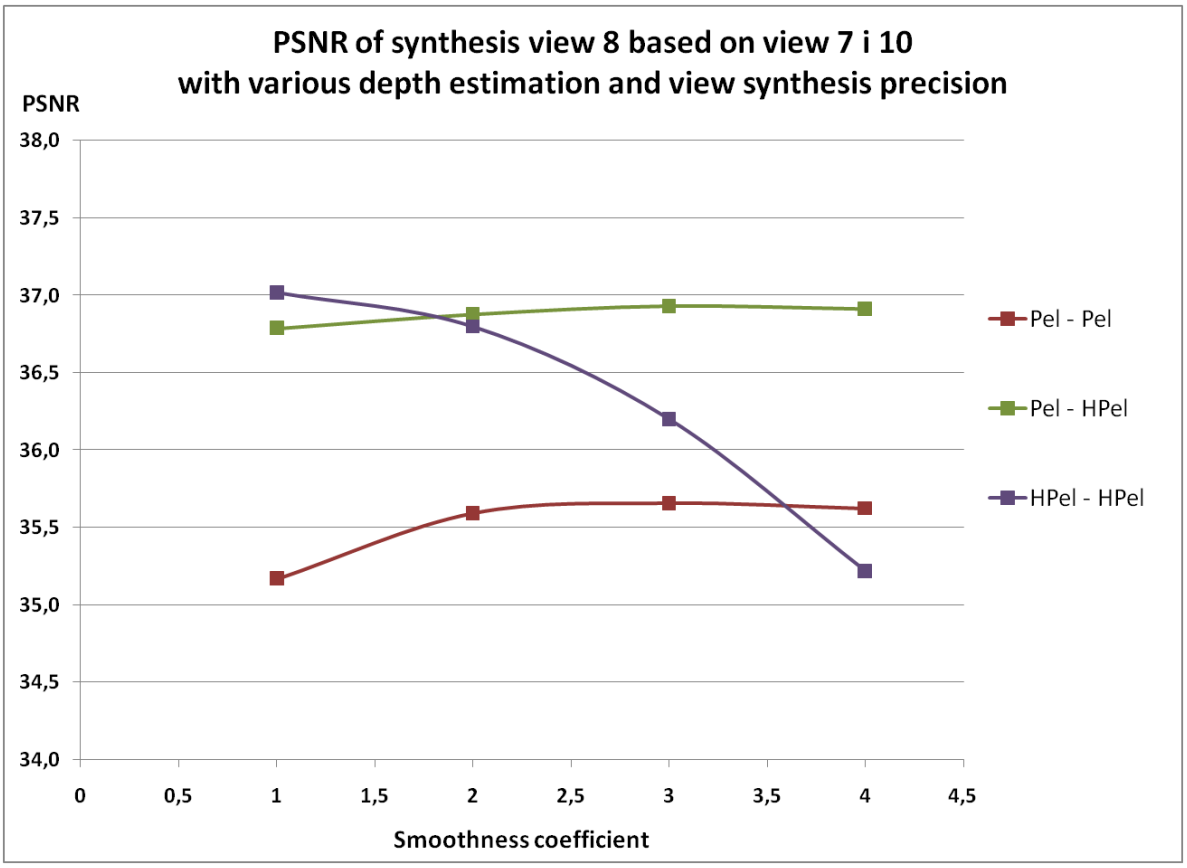

Figure 2. PSNR of synthesized virtual views (left and right) for 'Book arrival' sequence for various precision modes (depth estimation precision – view synthesis prediction).

For 'Book arrival' sequence, the best results can be attained with use of Pel-HPel mode (pixel precise depth estimation, Half-pixel view synthesis Table 2, Figure 2), and thus there is no need to use sub-pixel precise depth estimation.

Table 3. Averaged PSNR of synthesized virtual views (left and right) for ‘Lovebird 1’ sequence for various pixel precisions.

Performance [PSNR]		Depth Estimation Precision		
		Pel	Hpel	Qpel
View Synthesis Precision	Pel	28,1675	-	-
	Hpel	28,5054	28,3882	-
	Qpel	28,5104	28,4053	-

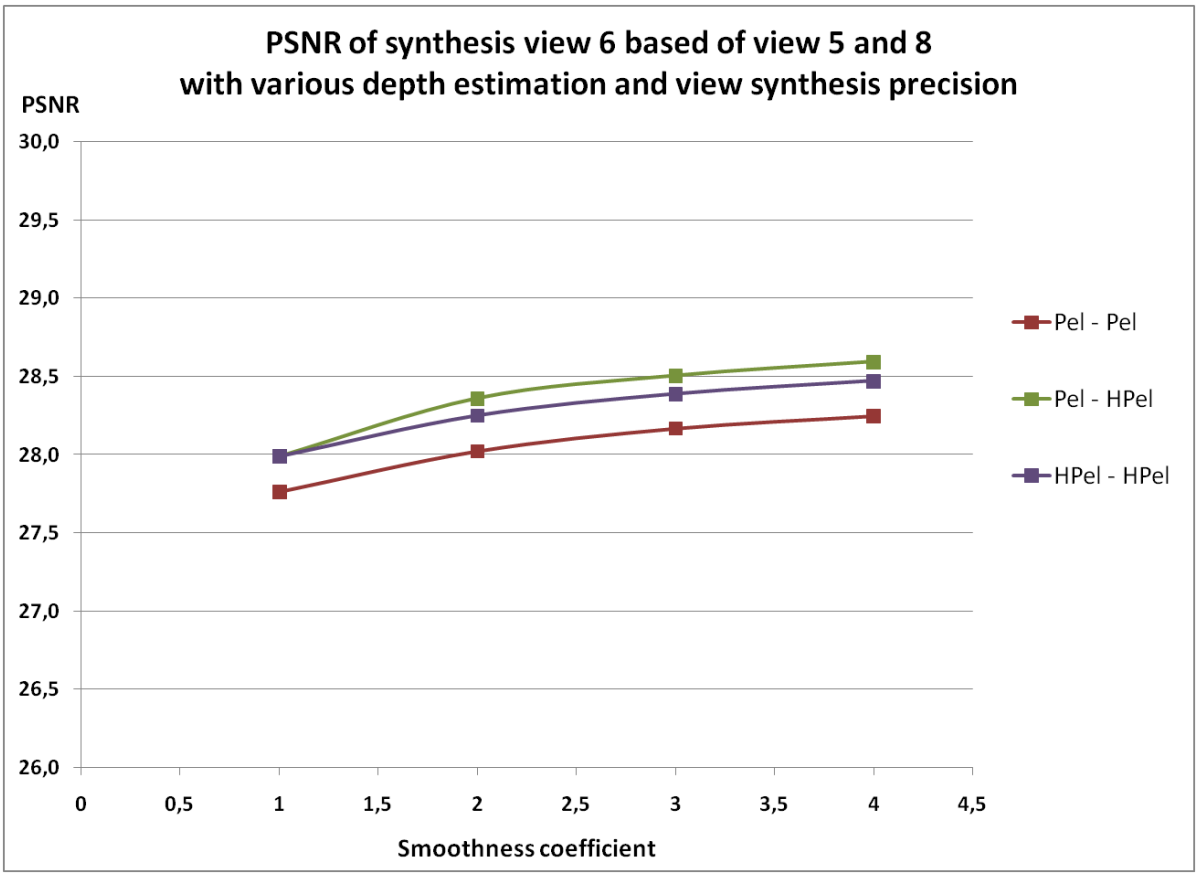

Figure 3. PSNR of synthesized virtual views (left and right) for ‘Lovebird 1’ sequence for various precision modes (depth estimation precision – view synthesis prediction).

In case of ‘Lovebird 1’ sequence, selection of depth estimation / view synthesis pixel precision modes makes almost no difference in performance of the tool-chain, however pixel precise depth estimation and half-pixel precise view synthesis pair outperforms other modes (Table 3, Figure 3)

Table 4. Averaged PSNR of synthesized virtual views (left and right) for 'Alt Moabit' sequence for various pixel precisions.

Performance [PSNR]		Depth Estimation Precision		
		Pel	Hpel	Qpel
View Synthesis Precision	Pel	34,0911	-	-
	Hpel	35,6875	35,3875	-
	Qpel	35,6817	35,3665	35,7263

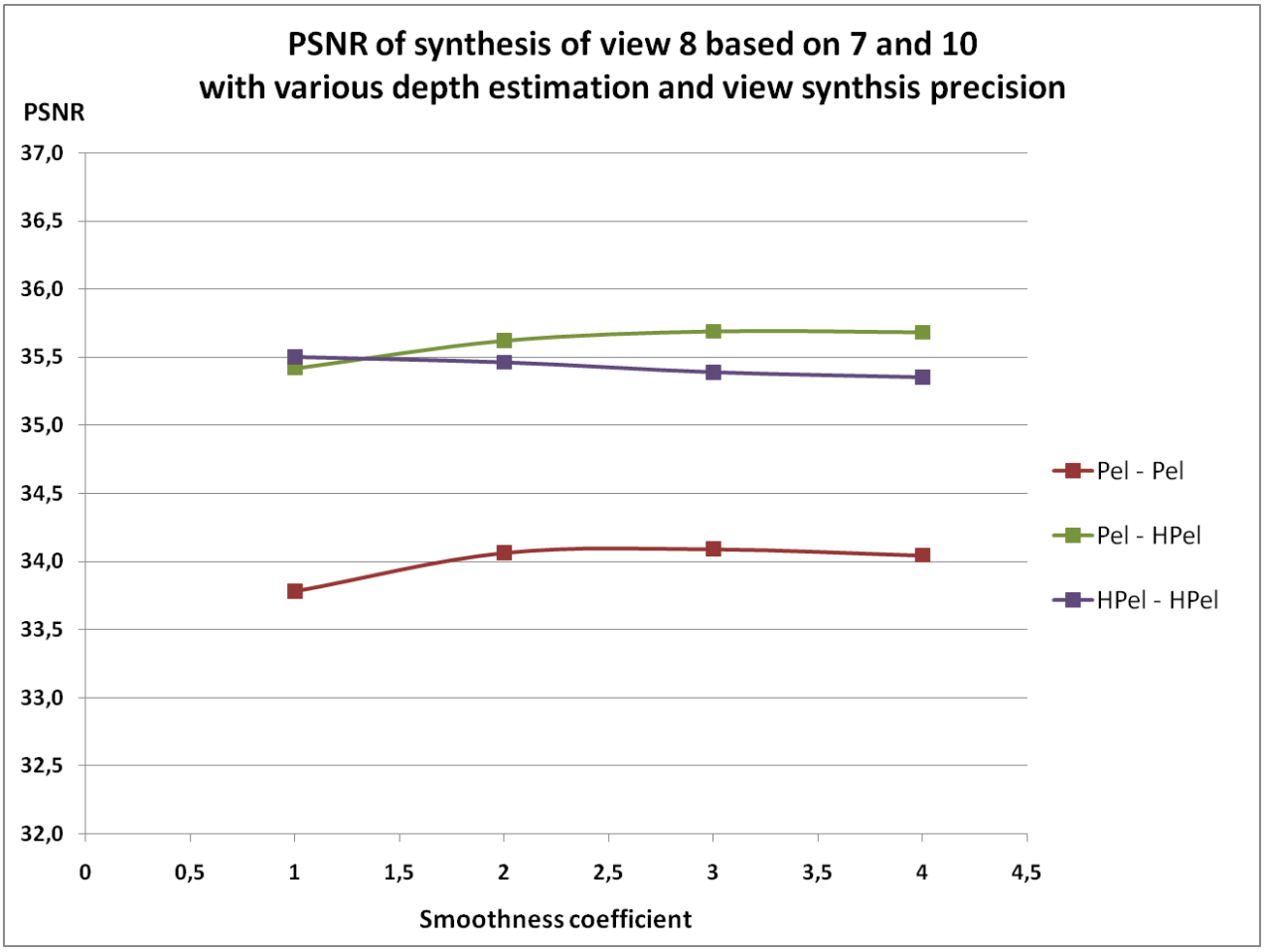

Figure 4. PSNR of synthesized virtual views (left and right) for 'Alt Moabit' sequence for various precision modes (depth estimation precision – view synthesis prediction).

As can be seen in Table 4 and Figure 4, pixel precise depth estimation performs alike sub-pixel depth estimation (half-pixel precise view synthesis in both cases) with little advantage for the first. Because pixel-precise depth estimation is slightly better and considerably less computationally expensive, it is recommended to use Pel-HPel mode.

Table 5. Averaged PSNR of synthesized virtual views (left and right) for ‘Dog’ sequence for various pixel precisions.

Performance [PSNR]		Depth Estimation Precision		
		Pel	Hpel	Qpel
View Synthesis Precision	Pel	25,0661	-	-
	Hpel	23,6088	29,3593	-
	Qpel	23,6088	28,3643	29,4145

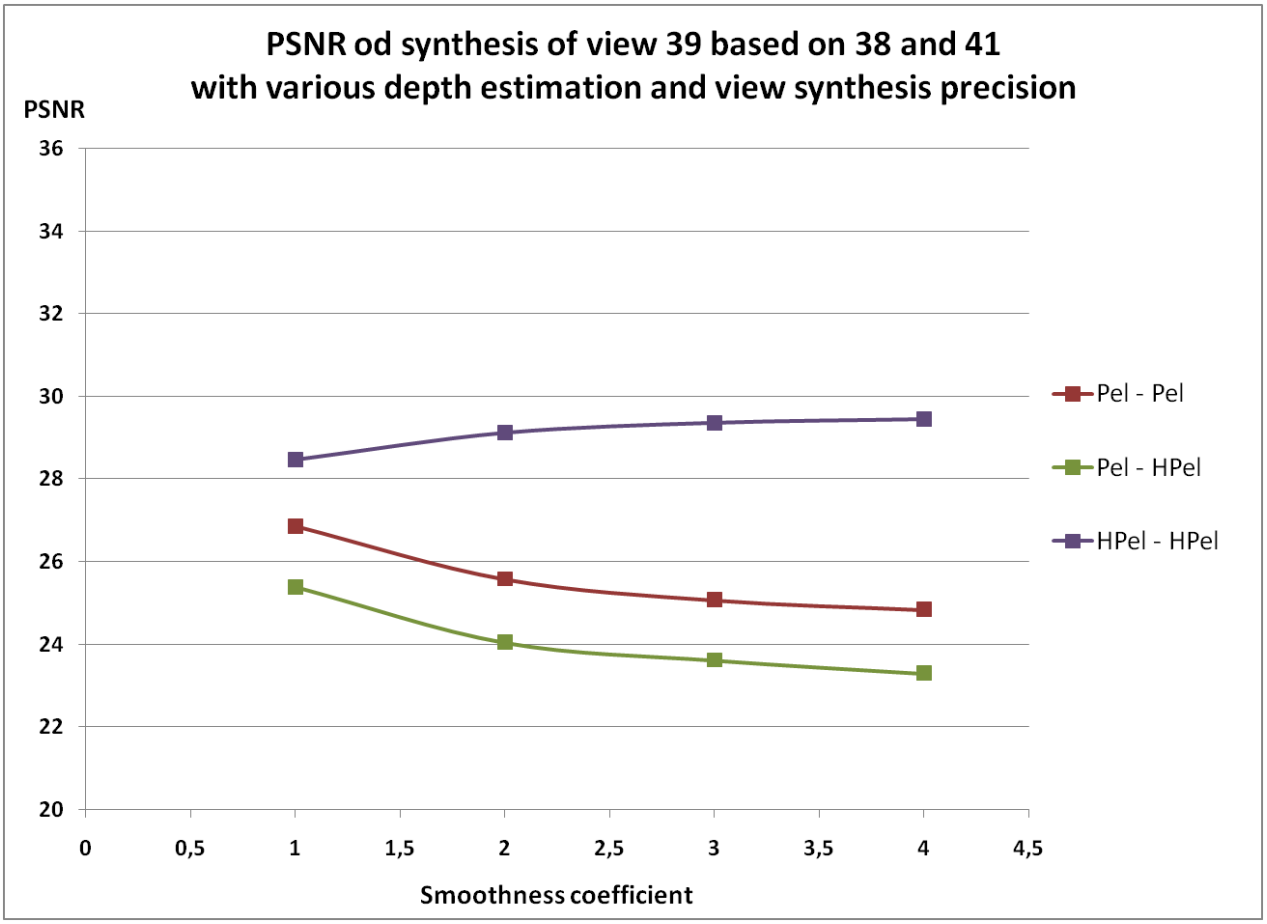

Figure 5. PSNR of synthesized virtual views (left and right) for ‘Dog’ sequence for various precision modes (depth estimation precision – view synthesis prediction).

In case of ‘dog’ sequence, HPel-HPel mode outperforms other modes in range allowed for smoothing coefficient (higher that 1.0). It is not known, whether Pel-Pel, Pel-HPel and HPel-HPel curves cross for lesser values of this coefficient.

4 Conclusions

- Pixel-precision mode in Depth Estimation Reference software has no impact on quality of synthesized views.

- The gain attained in sub-pixel precision Exploration Experiments comes from pixel-precision in View Synthesis Reference Software.

- It is recommended to limit computational power required by further experiments by use of pixel-precise depth estimation and sub-pixel precise view synthesis, or to use more efficient sub-pixel depth estimation technique.

5 References

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

[2] H. Kimata, A. Smolic, K. Müller, "AHG on 3D Video and FTV Coding" MPEG 2008/M15727, Busan, Korea, October 2008.

[3] "Description of Exploration Experiments in 3D Video Coding" MPEG 2008/W9991, Hannover, Germany, July 2008.