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

ISO/IEC JTC1/SC29/WG11
MPEG2008/M15832
October 2008, Busan, Korea

Source	**Poznań University of Technology**
Title	**3DV/FTV EE1 and EE2 results on Alt Moabit sequence**
Sub group	**Video**
Authors	**Krzysztof Wegner** (**kwegner@multimedia.edu.pl**) and **Olgierd Stankiewicz** (**ostank@multimedia.edu.pl**) Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland

1 Introduction

This document presents results of Exploration Experiments (EE1&EE2) performed on “Alt Moabit” sequence [2] and is in response to W9991 "Description of Exploration Experiments in 3D Video Coding" [1].

2 Experiment conditions

Experiments were performed basing on W9991 [1] guidelines, specifically Part 1 - narrow baseline case, which are as follows:

- Select stereo pair from data set, i.e. an original left view OL and an original right view OR (OL=8, OR=9)
- Estimate depth corresponding to neighboring original views NL (left) and NR (right) (NL=7, NR=10), using any available camera
- Synthesize views (synthesized left SL and synthesized right SR) at positions of OL and OR from NL+D and NR+D
- Bring synthesized video to the meeting
- Compare OL-OR with SL-SR subjectively

The test were performed on ‘Alt Moabit’ [2] sequence with following views selected as OL-OR and NL-NR.

Table 1. The specification of view for EE experiment.

Data set	OL-OR	NL-NR
Alt-Moabit	8-9	7-10

3 EE1 results

Tests performed as a part of Exploration Experiment part 1 were aimed at evaluation of quality of: matching methods, sub-pixel precision modes and temporal-consistency. All tests has been performed over wide range of Smoothing coefficient - from 1.0 to 6.0. Unfornatelly, lesser values of ‘smoothing’ coefficient that 1.0 are not allowed by the software. It can be estimated that values lesser that 1.0 would give better results.

3.1 Matching method selection

There are three matching methods available in Nagoya depth estimation reference software [3]: Conventional, Disparity-based and Homography-based. Test exploiting all of these modes were performed to find which is the most adequate , accurate and thus the best choice for further tests.

The results (presented in Figure 1 and Figure 2) show that the matching method has very limited impact on performance of the algorithm (diffirences less than 0,01 dB) and thus Conventional matching method was selected for all further experiments, because it is the fastest one.

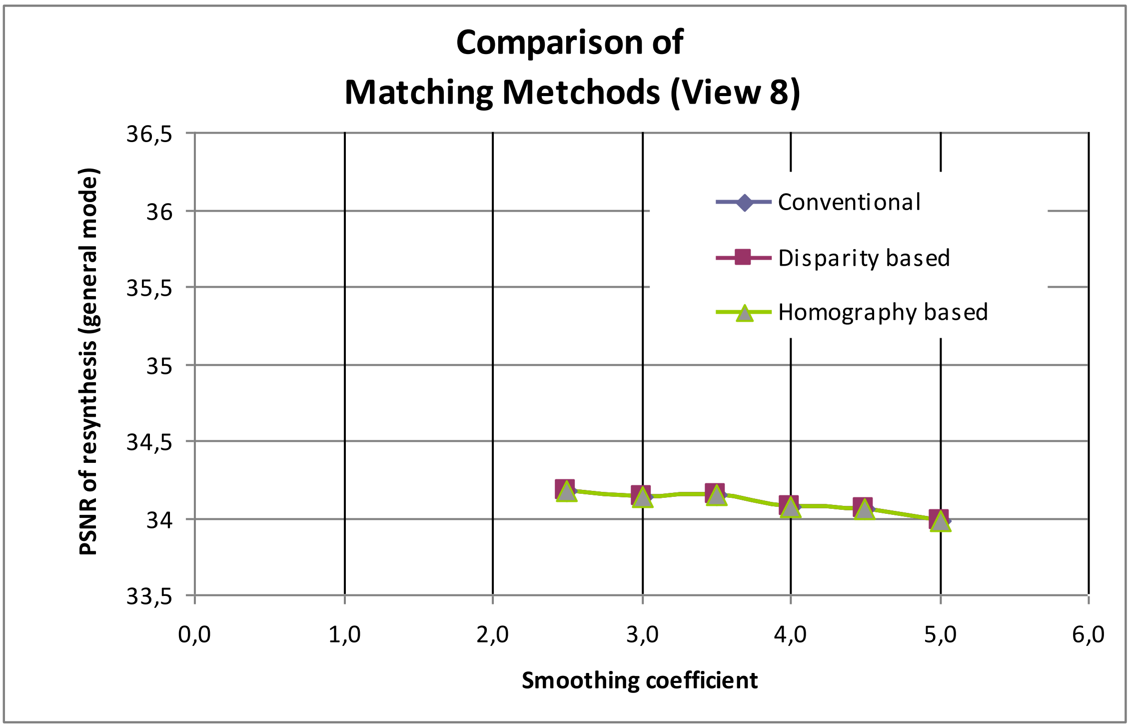

Figure 1. Comparison of matching methods in depth estimation software with respect to PSNR of resynthesis (reference software) of View 8.

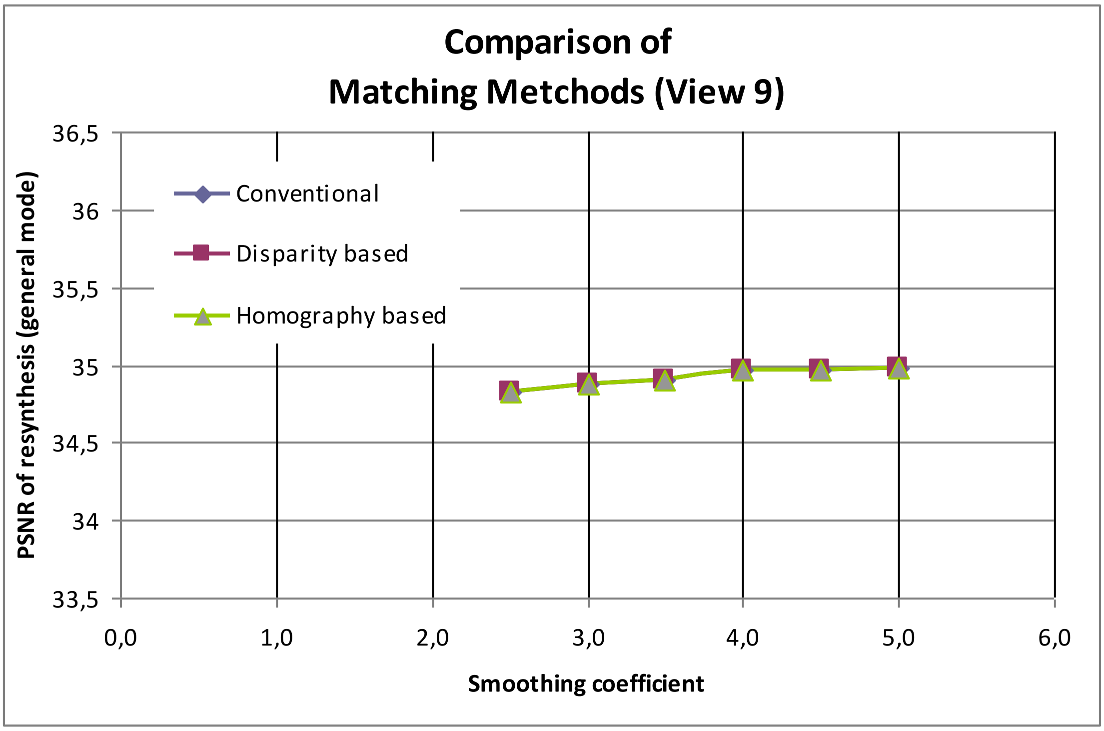

Figure 2. Comparison of matching methods in depth estimation software with respect to PSNR of resynthesis (reference software) of View 9.

3.2 Sub-pixel tests results

The tests have been performed with three degrees of pixel-precision: full-pixel (the least accurate), half-pixel and quarter-pixel (the most accurate).

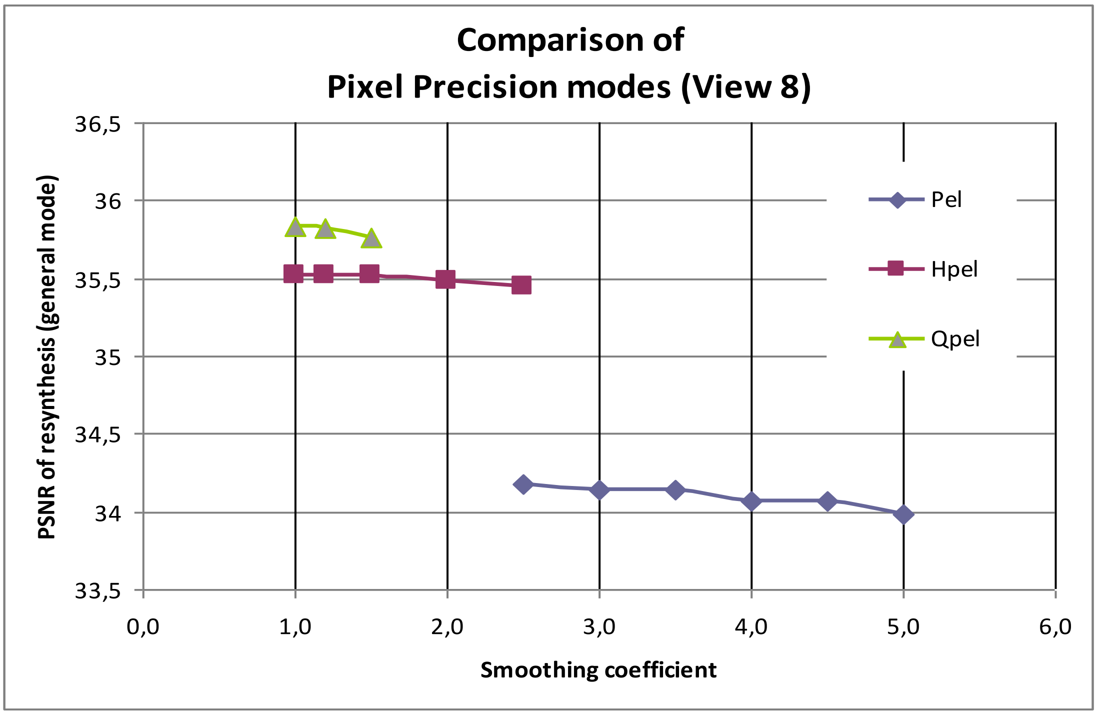

Figure 3. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (reference - general mode) of View 8.

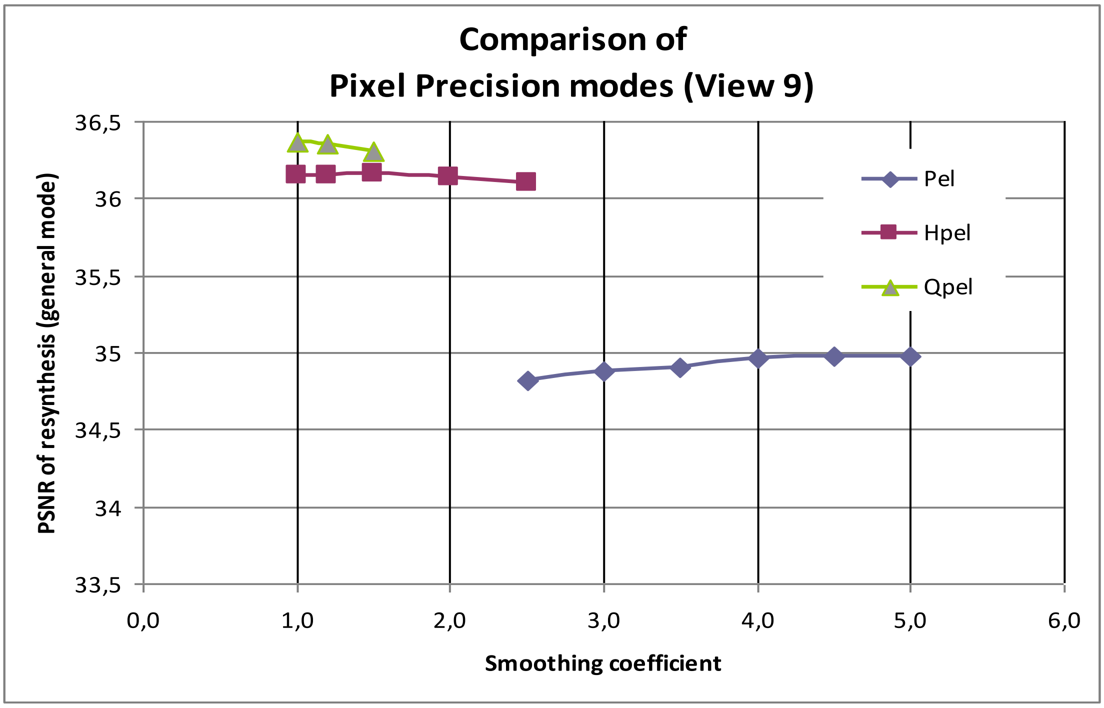

Figure 4. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (reference - general mode) of View 9.

Nagoya reference resynthesis software allows for use of two resynthesis modes: general mode and 1-D mode (for special parallel cases). Figures 5 and 6, when compared to Figures 3 and 4 respectively, show that 1-D mode is less efficient (about 0,25 dB of loss) than the general mode.

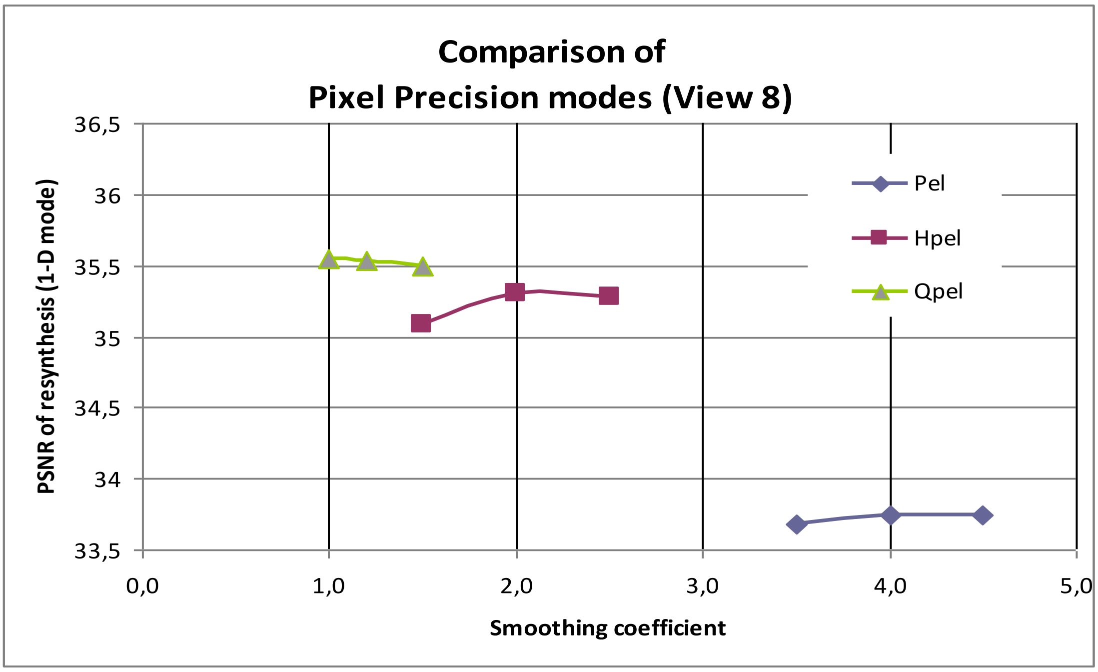

Figure 5. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (reference - 1-D mode) of View 8.

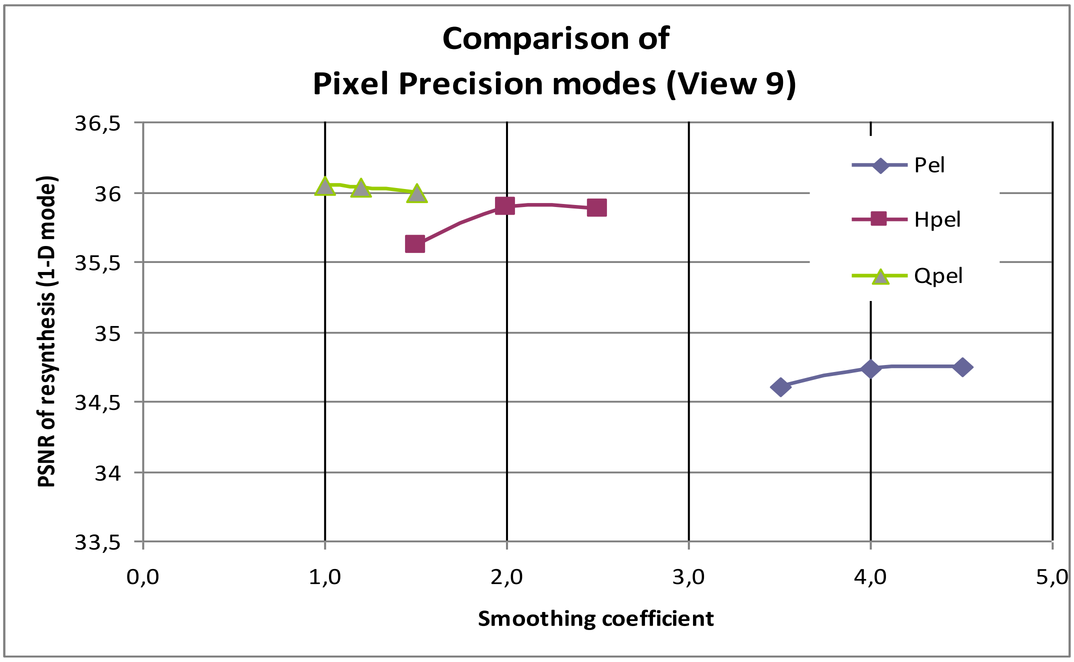

Figure 6. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (reference - 1-D mode) of View 9.

The overall results (presented in Figures 3-6 for views 8 and 9) show that the increase of precision gives considerable gain, yet this gain reaches its plateau near quarter-pel precision. Increase of precision from half-pixel to quarter-pixel gives increase of PSNR of about 0.3dB, and it can be estimated that further increase of precision (to 1/8th pixel) would give even lesser gain (of approximately 0.1dB) – these measures and estimation are shown in Figure. 7.

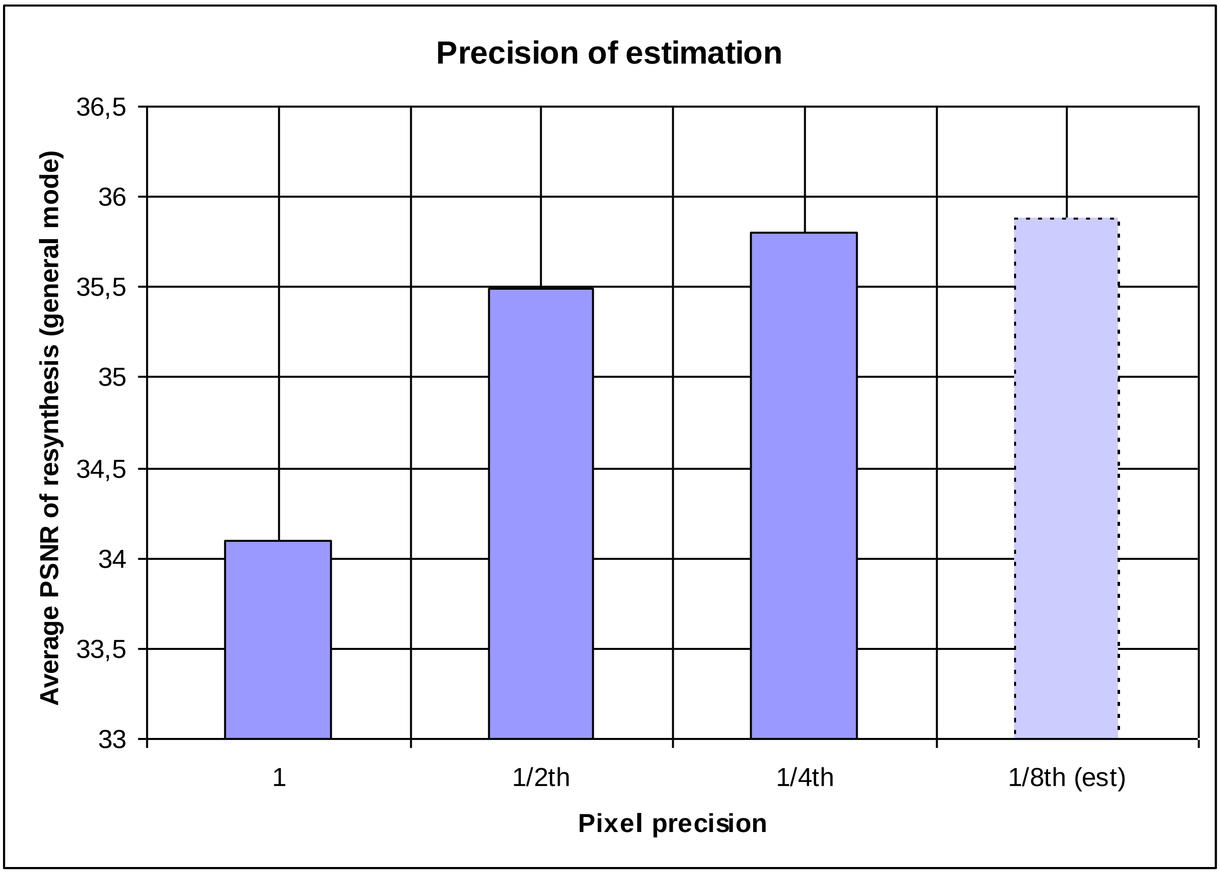

Figure 7. Average resynthesis quality by PSNR versus tested pixel precision (1-1/4th pixel) and estimated (1/8th-pixel).

3.3 *Temporal Consistency modification results*

The Temporal Consistency modification for reference software has been provided by GIST [4]. The amount of temporal-consistency introduced by the software is controlled by 'slope of the weight function' (slope) and distance between key-frames. The range of 0.0 (no temporal-consistency) to 2.0 has been tested and length of 5, 10 and 15 frames between key-frames.

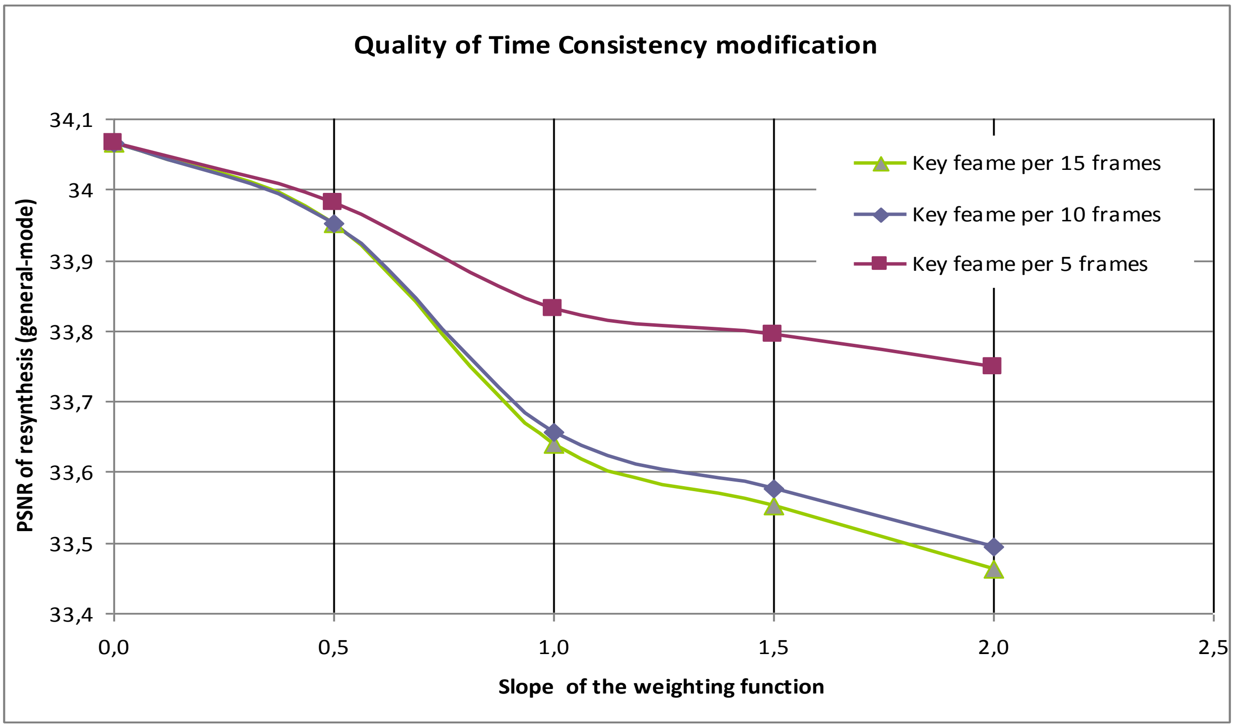

Figure 8. Quality of resynthesis for temporal-consistent depth-map depending on various Key frame and 'slope' parameter settings.

The results show that there is a decrease of objective quality when Temporal Consistency modification is used. The difference for 'slope'=2 is of about 0,3dB when key frame is placed every 5 frames, 0,6dB when key frame is placed every 10 regular frame, and of about 0,8dB when key frame is placed every 15 frames.

3.4 *Conclusions*

- There are big fluctuations of quality depending on synthesized view selection (view 8 or 9), up to about 0.5dB of PSNR.
- Selection of matching method (Conventional, Disparity-based or Homography-based) in Nagoya depth estimation software has negligible impact on quality and resynthesis results
- Sub-pixel (Half and Quarter-pixel) estimation of depth map yields with considerable gain of PSNR.
- There is no evidence that further increase (than quarter-pixel) of pixel-precision would bring reasonable gain.
- Simplified method of resynthesis in Nagoya reference software (1-D mode) gives worse results than the general one, both objectively and subjectively.
- Increase of key-frame distance in Temporal Consistency modification tests leads to decrease of visual quality both in manner of PSNR and subjective experiences.

- In general, “motion blur” is too high with use of Temporal Consistency modification, and probably other methods are required.

4 EE2 Results

Tests performed as a part of Exploration Experiment part 2 were aimed at evaluation of quality of ViSBD resynthesis software provided by Thompson [5].

4.1 ViSBD

As in case of Nagoya reference software, synthesis results have been tested against various values of ‘smoothing’ coefficients and three pixel precision modes. ViSBD software also provides two modes of operation: with splatting and without splatting. Results (Figures 9-12) show that turning splatting “off” is more efficient than turning “on”. The difference is of about 0.8dB.
Slicing “off” option, as the better one, has been selected for further comparisons (with Nagoya reference software). The results are shown in Figure 13. It can be noticed that ViSBD performs worse than reference software. The difference varies from 1dB (Pixel- precision) to almost 2dB (half-pixel precision).

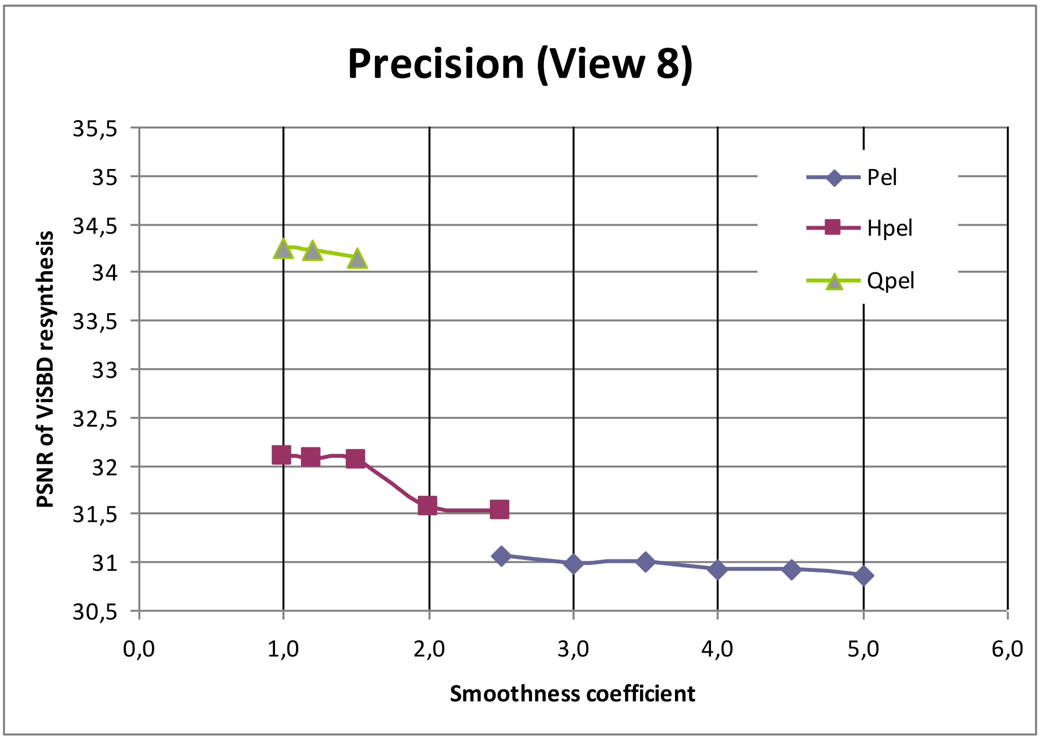

Figure 9. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (ViSBD, splatting “on”) of View 8.

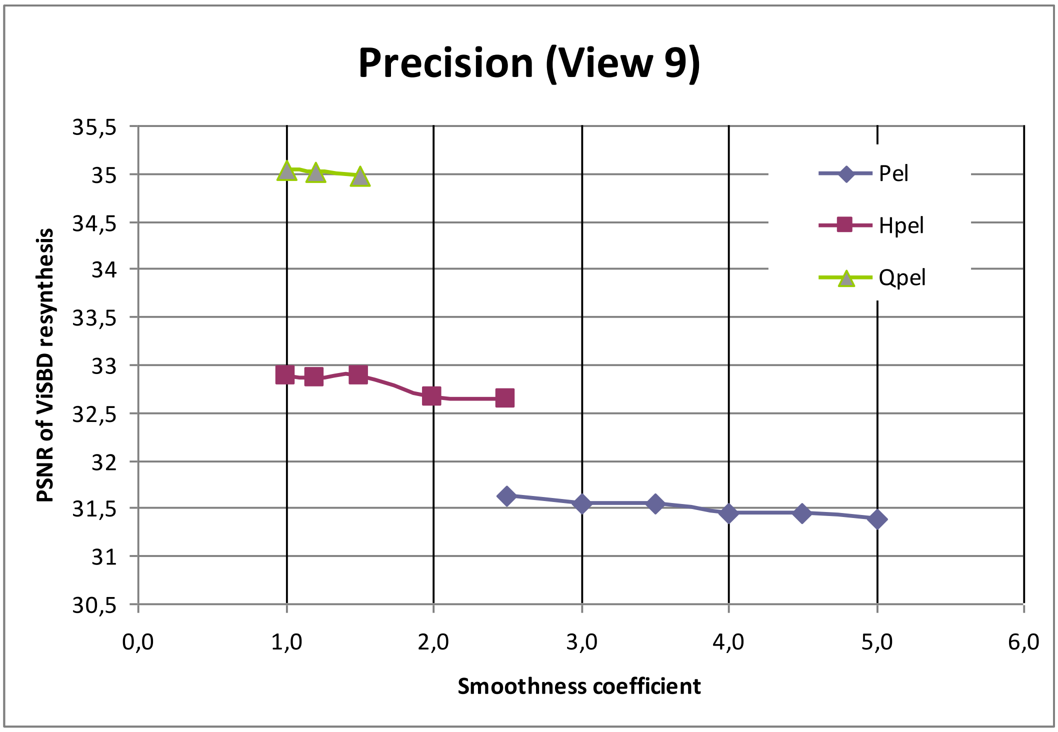

Figure 10. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (ViSBD, splatting “on”) of View 9.

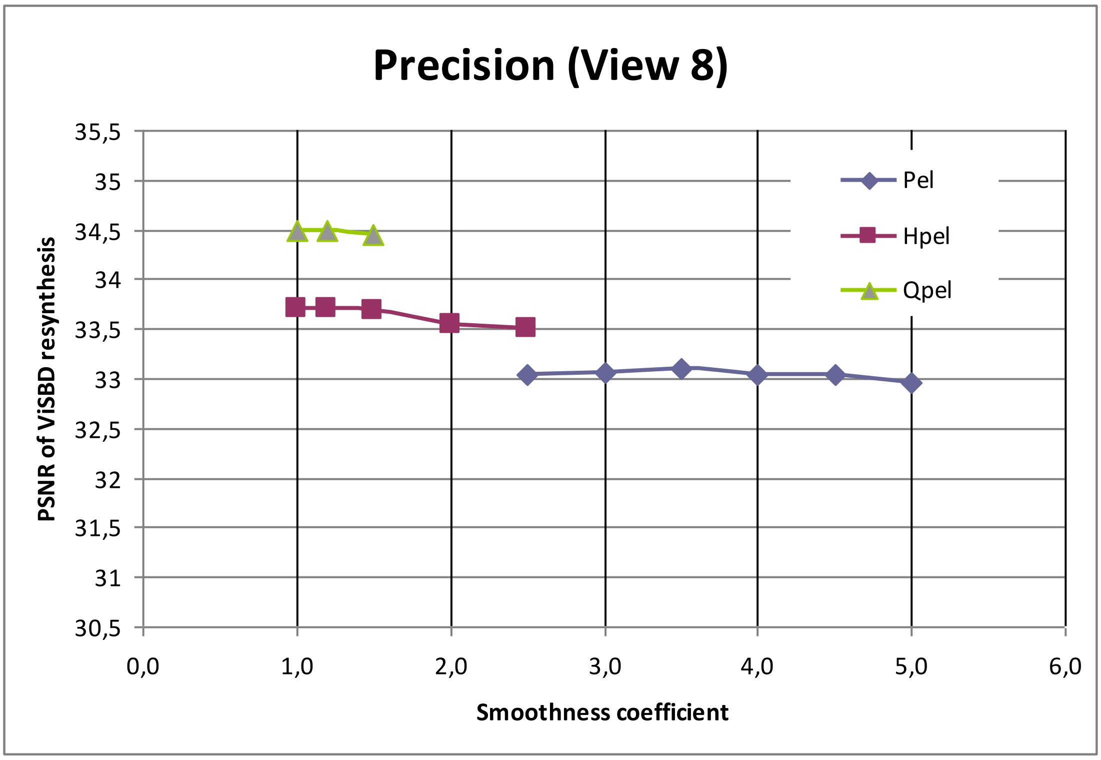

Figure 11. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (ViSBD, splatting “off”) of View 8.

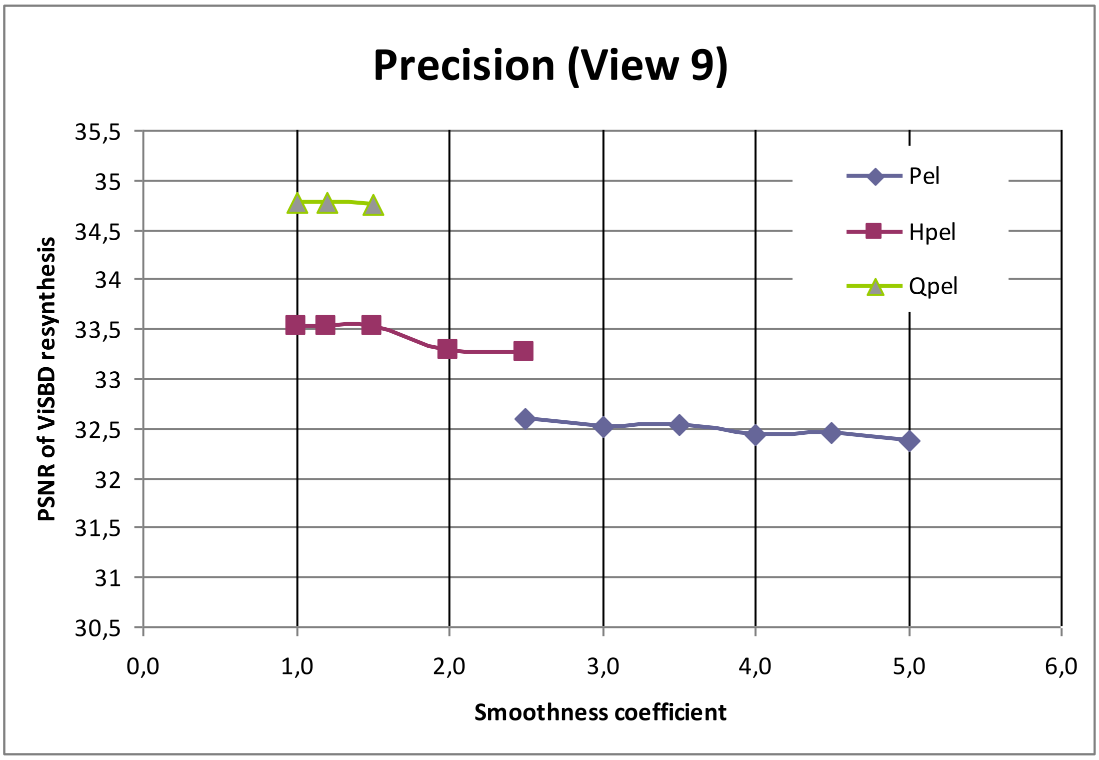

Figure 12. Comparison of matching precision in depth estimation software: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) with respect to resynthesis (ViSBD, splatting “off”) of View 9.

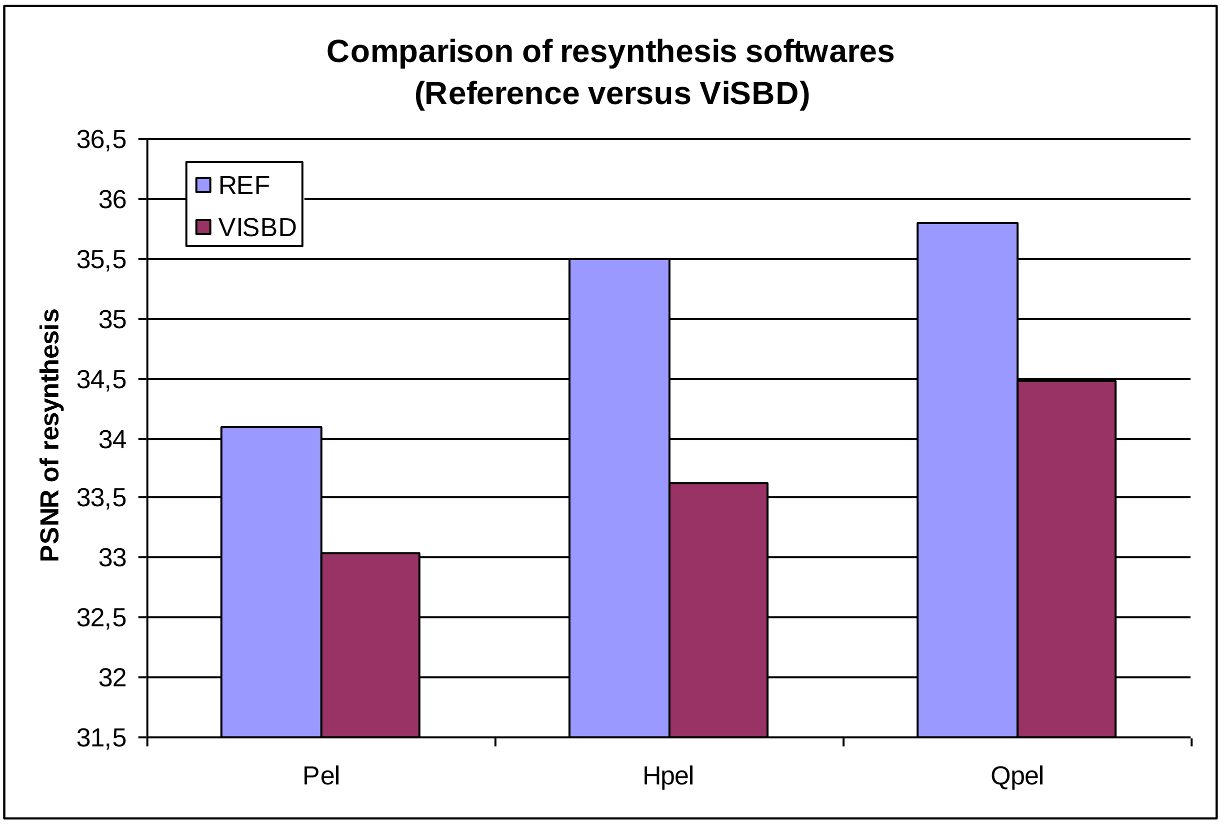

Figure 13. Comparison of ViSBD resynthesis software (splatting “off”) with Nagoya reference software with respect to PSNR for: full-pixel (Pel), half-pixel (Hpel) and quarter-pixel (Qpel) accuracy of depth maps.

4.2 Conclusions

- There are big fluctuations of quality depending on synthesized view selection (view 8 or 9), up to about 0.5dB of PSNR.
- ViSBD gives worse resynthesis results than Nagoya reference software (1÷2dB), and slightly worse subjective quality.
- 'Splatting' option has little impact on results quality ("Splatting off" is superior to "splatting on" about 0.4dB).

5 Subjective quality evaluation

For evaluation of aubjective quality we have used the idea of Mean Opinion Score (MOS). The MOS (in case of our study) is expressed by a 5-point continuous scale. Rating the quality range from 1 ("very bad with annoying impairments/artifacts") to 5 ("imperceptible"). The reference sequence was a'priori unknown and order of appearance of synthesized sequences has been randomly chosen. The test has been carried out on the group of 15 human subjects.

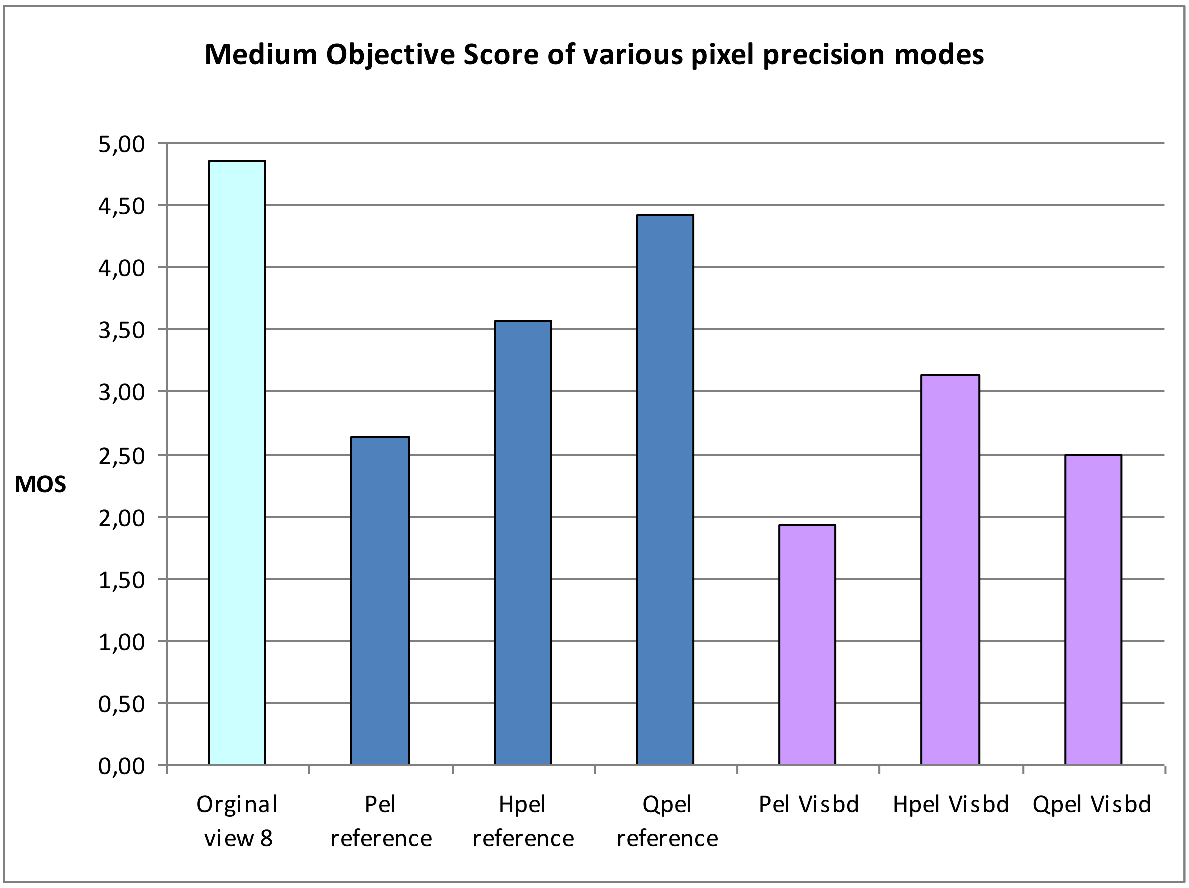

Figure 14. Comparison of various pixel precision modes of depth map estimation by resynthesis evaluation for both Reference Software and ViSBD.

As can be seen in the Figure 14, the best performing option is use of quarte-pixel precision and Nagoya resynthesis software – it is just about 0.5 point below original view (with rating of 4.86). In general, ViSBD gives subjectively worse result that the reference software, rating about 1 point lower.

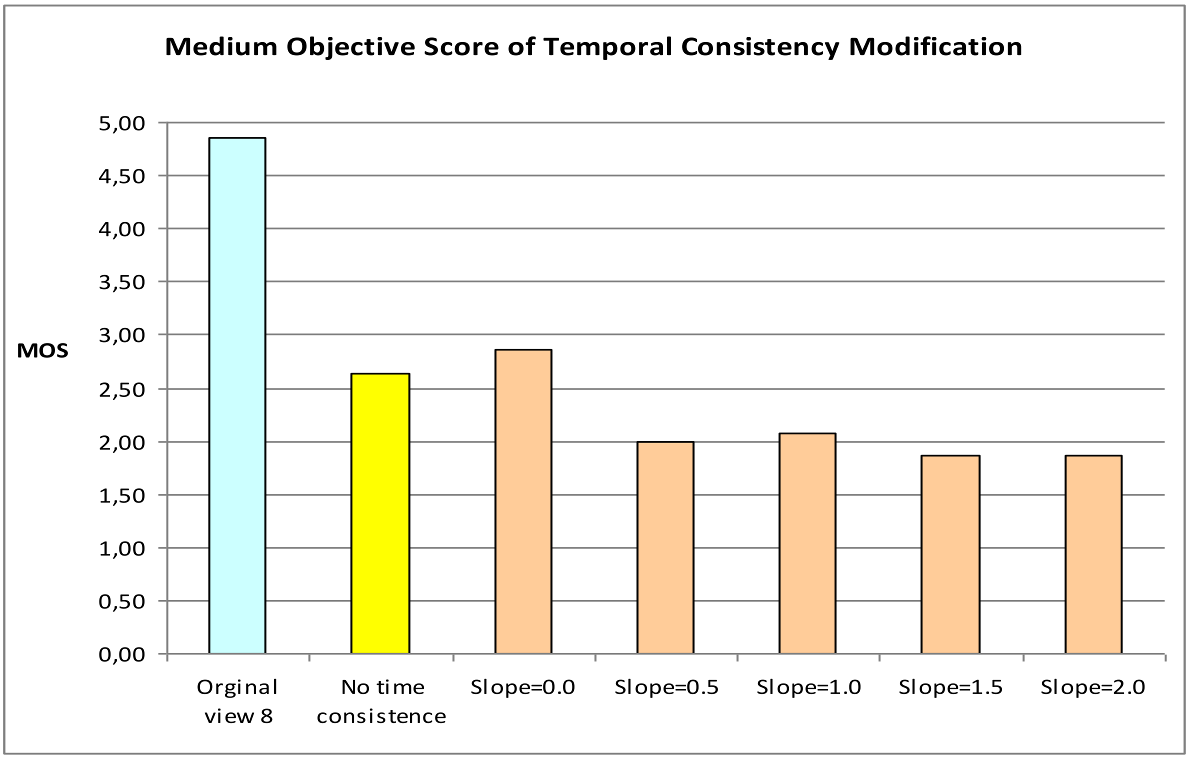

Figure 15. Comparison of various 'slope' parameters.

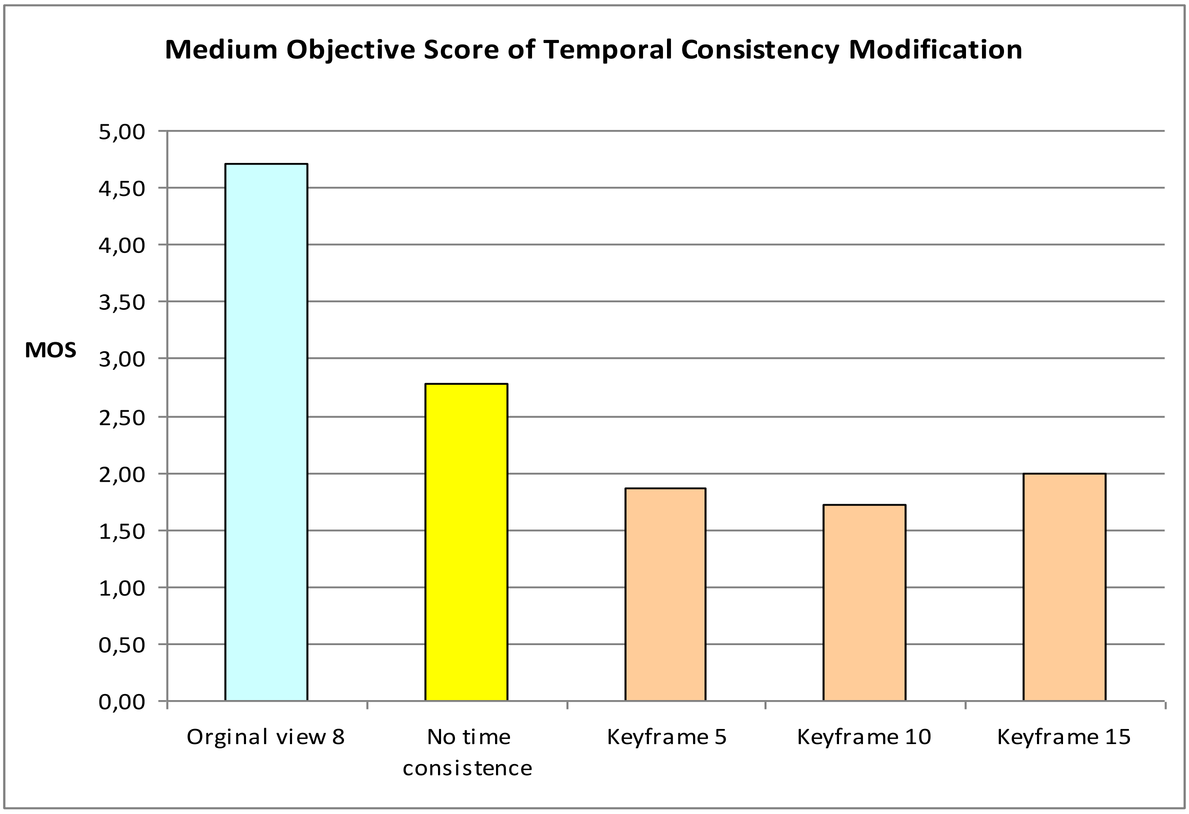

Figure 16. Comparison of various key-frame distances.

Figures 15 and 16 show that Temporal Consistency modification introduces loss of subjective quality expressed as loss of about 2.0 points. In particular, one can notice that neither change of ‘slope’ parameter nor distance between key-frames lead to significant changes of perceptive quality.

6 References

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

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

[3] M. Tanimoto, T. Fujii and K. Suzuki, “Multi-view depth map of Rena and Akko & Kayo”, ISO/IEC JTC1/SC29/WG11, M14888, October 2007.

[4] S. Lee and Y. Ho “Enhancement of Temporal Consistency for Multi-view Depth Map Estimation” MPEG 2008/M15594, Hannover, Germany, July 2008.

[5] D. Tian, P. Pandit, C. Gomila “Simple View Synthesis” MPEG 2008/M15696, Hannover, Germany, July 2008.