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/M17612
April 2010, Dresden, Germany

Title	**Estimation of temporally consistent depth maps using noise removal from video**
Sub group	**Video**
Authors	**Olgierd Stankiewicz** (ostank@multimedia.edu.pl), **Krzysztof Wegner** (kwegner@multimedia.edu.pl) and **Marek Domański** Poznań University of Technology, Chair of Multimedia Telecommunications and Microelectronics, Poznań, Poland

1 Introduction

This document presents a new tool for providing temporally consistent depth map. This tool is an alternative for Temporal Consistency tool in DERS [1]. As shown, the results are very promising: depth maps produced with use of proposed technique are very stable. Depth flickering, which typically results in annoying artifacts in synthesis is almost entirely removed for static objects.

2 Main Idea

We propose a novel approach to problem of temporal consistency. To tackle temporal inconsistency we propose to eliminate its cause. Depth map fluctuations are caused by noise, mainly temporal. We propose to employ noise reduction on video before depth estimation. Each view of a multi-view video sequence, is independently denoised in time and then feed to a depth estimation algorithm (Fig. 1). As we show later, depth maps obtained in such way are more consistent in time.

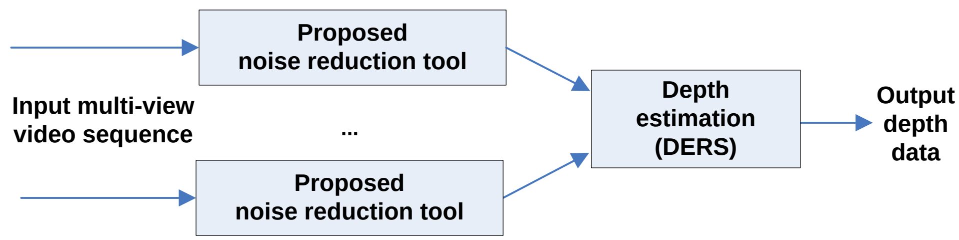

Figure 1. Idea of the proposed approach.

For noise reduction we use our simple denoising algorithm. Our noise reduction technique employs temporal filtering and focuses on regions where it applies best – to steady regions of the sequence. These are the regions, where the most of inconsistency in depth data occurs.

3 Noise reduction technique

Our method consists of three main phases (Fig. 2):

- Motion detection, where pixels are classified as moving or steady,
- Noise filtering, where steady pixels are filtered in time, and
- Artifact removal, where errors of motion detection stage are repaired.

Moving pixels are left unchanged during the entire processing. This is motivated by fact that there is uncertainty of whether motion cues (generated by Motion Detector) are caused by noise or by motion itself. Moreover, temporal filtering applies best to steady regions because in such case there is no need for computationally consumptive motion estimation and compensation.

All phases of the algorithm are performed in three pipelines of frames (Figure 2):

- original frames (input of the algorithm),
- binary motion maps (by-product of the algorithm),
- denoised frames (result of the algorithm).

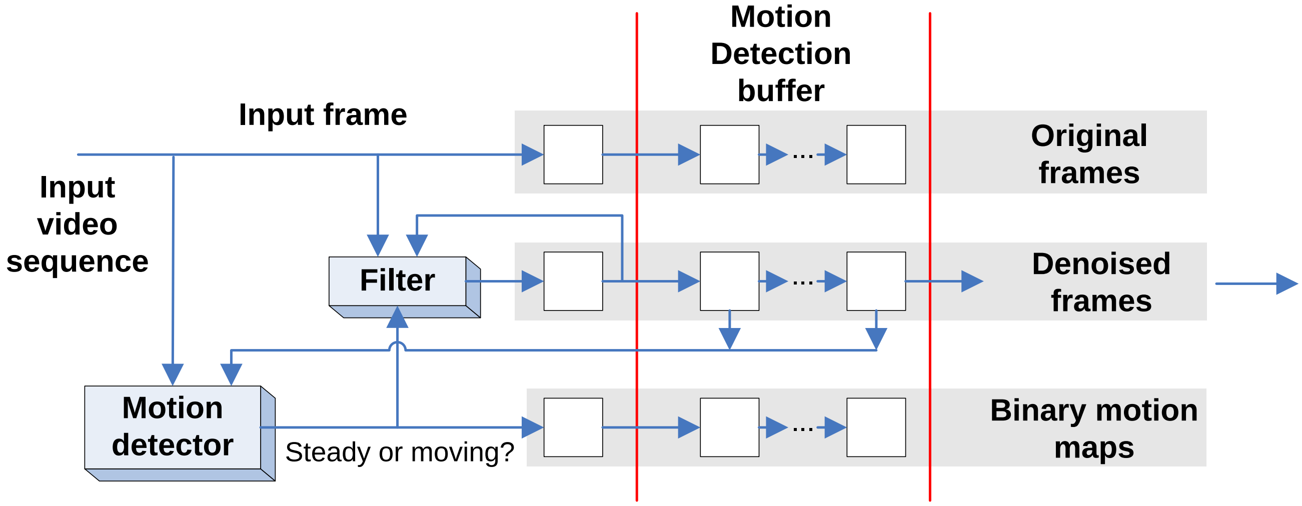

Figure 2. Block scheme of the algorithm.

The role of Motion Detector is to classify pixels from input frame as moving or as steady. Result of this classification is combined into a binary map called by us motion map (Figure 3). This map is generated with use of some non-linear filtering operations that use rectangular masks of 9x9 pixels.

Figure 3. Motion map (right) obtained for exemplary frame (left)
(white pixels – moving, black pixels – steady).

Filtering block performs the main denoising operation. As mentioned before, pixels classified as moving are left unchanged and are not modified by the algorithm. Pixels classified as steady are independently filtered in time with respect to previously filtered frame, stored in "Denoised" pipeline of the algorithm (Fig. 2). The idea behind this is to "freeze" the noise on steady pixels, so that the depth estimation is not confused with fast varying pixel values. In our work we have exploited simple low-pass IIR (Infinite Impulse Response) filter of first-order. Low order filter was chosen to reduce computational complexity and to allow slight changes in the scene (e.g. day-time lighting).

4 Experimental results

To assess influence of presented noise reduction technique on quality of depth data, depth maps have been estimated from "denoised" and from "original" video with use of the latest version of DERS [1]. Temporal Consistency mode in DERS was **turned off**. A standard EE-like evaluation procedure , based on view synthesis has been performed. Some test sequences [1,2] have been used, for both objective (PSNR) and subjective (Mean Opinion Score – MOS) evaluation.

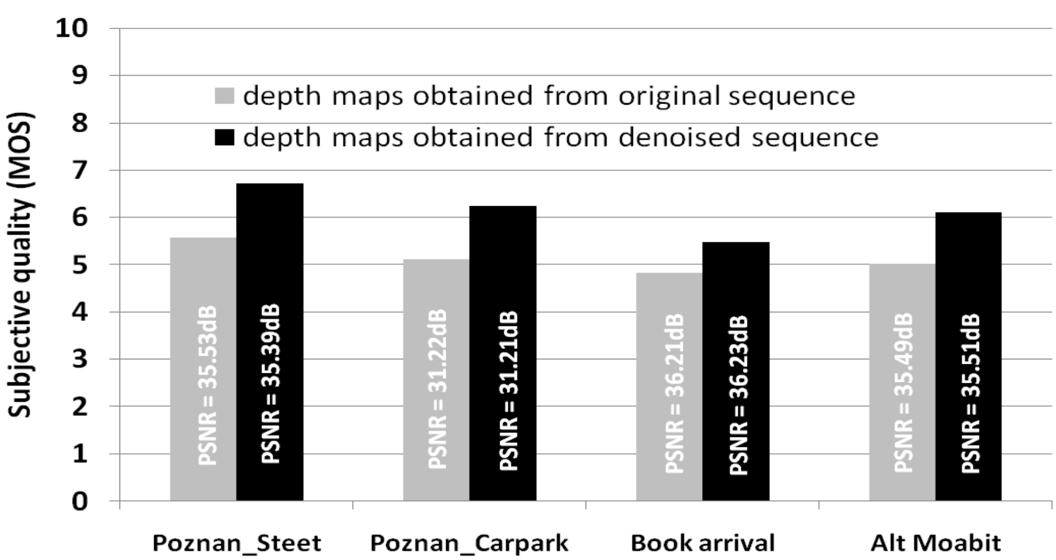

Figure 4. Depth map quality assessment by assessment of quality of synthesized view.

Results presented in Figure 4 show that use of proposed approach increased subjective quality of synthesized views from about 0.7 to 1.2 MOS points. It can also be noticed that PSNR levels have not changed. The latter is not surprising, because PSNR measure is not designed to asses quality of temporal consistency, and because only original sequences have been used for synthesis.

Figure 4. Exemplary results of proposed technique: original (left) and denoised (right).
a) image   b,c) depth maps for two consecutive frames   d) difference between depth maps.

Figure 4 show exemplary results attained with and without use of proposed approach. As can be noticed on Fig. 4a, Moving objects (people) are left unchanged while background (wall and cars) is significantly denoised. It is worth to notice that denoised images are not blurred, because only temporal filtering is employed. Although quality of depth maps (Fig. 4b,c) has not changed, temporal consistency expressed as difference between frames (Fig. 4d) is vastly improved. As shown, background remains static (black means no changes) and thus is consistent is time. Of course, there is no improvement over moving objects, because they are not filtered.

5 Conclusions and recommendations

Conclussions:

- We have proposed a new tool aimed at providing temporally consistent depth maps,
- The technique is an alternative for Temporal Consistency in DERS, but is not exclusive,
- The subjective gains are considerable, and range from about 0.7 to 1.2 in MOS (1-10) scale.

Recommendations:

- To incorporate this new tool into DERS as an additional Temporal Consistency mode.

6 References

[1] M. Tanimoto, T. Fujii, M. P. Tehrani, M. Wildeboer „Depth Estimation Reference Software (DERS) 4.0”, ISO/IEC JTC1/SC29/WG11, MPEG 2008/M16605, London, UK, June 2009.

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