Generation of temporally consistent depth maps using noise removal from video

Abstract. This paper presents a novel approach for providing depth maps that are temporally consistent. Temporal consistency is attained by noise removal from video. Presented approach was evaluated with use of a simple noise reduction technique and state-of-the-are depth estimation algorithm. Experiments on standard multi-view test video sequences have been performed and yielded both subjective and objective results. These results provide evidence that the proposed approach increase temporal consistency of estimated depth maps.

Keywords: depth map estimation, temporal consistency, temporal noise removal.

1 Introduction

Depth map estimation is a technology that provides 3D representation of the scene [1]. Common approach to obtaining depth data is algorithmic estimation from video. Although many such algorithms are known in literature [2], depth estimation is still a challenge, even for the most advanced state-of-the-art techniques. One of the biggest of challenges in this research area, is how to provide depth maps that are consistent in time.

Typically, depth data for video is estimated independently for each frame of the sequence. Unfortunately, estimation that is independent in time, causes depth of objects in the scene to fluctuate, due to noise. Such fluctuations are adverse, because they lead to occurrence of artificial movement in 3D representation. Desired 'depth map temporal consistency' means that changes of the depth of objects in time are correlated with actual motion of the objects and do not vary from frame to frame in a random way.

Majority of state-of-the-art techniques that tackle temporal consistency, in various ways expand depth estimation algorithms into time domain. For example, in [3] authors propose to extend standard 4-neighborhood belief propagation depth map estimation scheme [4] to 6-neighborhood scheme by addition of temporal neighbors: from previous and from next frame. These neighbors are obtained by motion estimation. Therefore, depth value is optimized with respect to depth value in neighboring frames. In turn, authors of [5] propose segment-based approach. In order to provide temporally consistent depth value, apart from traditionally used spatial matching of segments, also temporal segment matching is performed. Such approach increase complexity of the whole depth estimation process, which already is computationally expensive.

We propose a novel approach to problem of temporal consistency. To tackle temporal inconsistency we propose to eliminate its cause. As mentioned before, depth map fluctuations are caused by noise, mainly temporal. We propose to employ noise reduction on video before depth estimation. As we show later, depth maps obtained in such way are more consistent in time.

Noise reduction is a well-known and widely recognized technical field. Wide variety of examples of noise reduction techniques can be found in [6,7,8]. Classical noise reduction techniques aim to provide a denoised image directly to the audience. In case of depth estimation, more artifacts are allowed, because denoised version of the image is only to be used for depth estimation. Thus, wider range of techniques can be considered.

Presented approach is a new general idea, because it can be applied to any depth estimation technique and any noise reduction technique without any modifications. Although that, for the sake of this paper we chose a fixed setup of these algorithms. We use depth estimation algorithm [9] that is being used as a reference for standardization of 3D television by ISO/IEC MPEG group. For noise reduction we use our simple denoising algorithm, presented in Section 3.

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.

2 Idea of the paper

This paper aims at production of better and more temporally consistent depth maps. The main idea consists in application of a temporal noise reduction technique before depth estimation algorithm. Each view of a multi-view video sequence, is independently denoised in time and then feed to a depth estimation algorithm (Fig. 1).

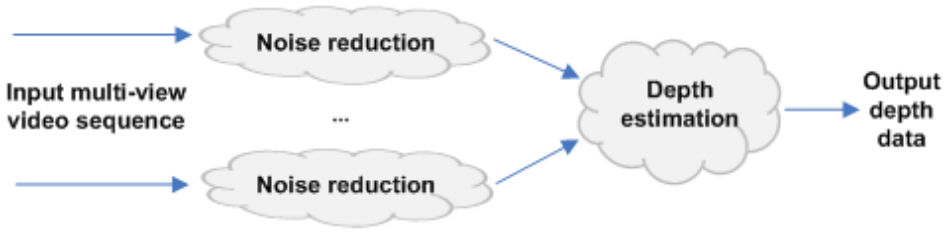

Fig. 1. Idea of the proposed approach. Any noise reduction and any depth estimation techniques can be used.

3 Noise reduction technique

For evaluation of the presented approach, we have employd a simple temporal 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, which we prefer to avoid.

All phases of the algorithm are performed in three pipelines of frames: original frames (input of the algorithm), binary motion maps (by-product of the algorithm) and denoised frames (result of the algorithm).

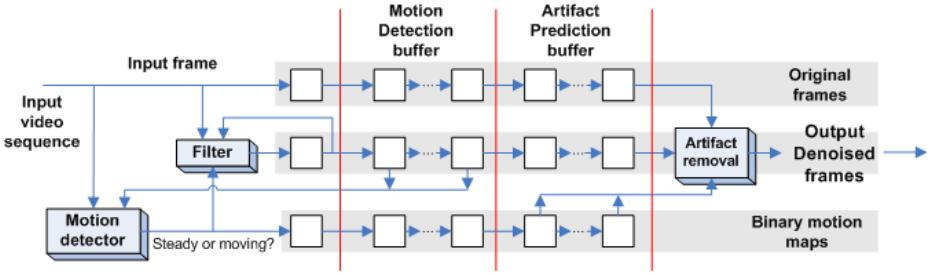

Fig. 2. Block scheme of the algorithm.

3.1 Motion Detection

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.

Our motion detection algorithm is presented in Fig. 3. Each pixel of input frame is compared with corresponding pixels of N previous frames, stored in Motion Detection Buffer, by means of absolute differences. These absolute differences are then maximized between frames and over RGB color components. Resulting maximum absolute differences are feed to two parallel paths: top and bottom (Fig. 3). These paths, give cues about motion that occur in neighboring pixels, for each pixel. Top path give map of cues about maximal motion, and bottom path gives map of cues about distributed motion.

Maximal motion cue map (top path) is obtained with use of dilation filter. Dilation is performed in rectangular window (of size 9×9 in experiments).

Distributed motion cue map (bottom path) is obtained by counting of pixels that exceed certain level (binarization) in window surrounding each pixel (window size is the same as in the top path).

Output of the motion detection, binary motion map (Fig. 4), is produced by combining of motion cue maps from top and bottom path. Pixel is marked as moving (white) if any of motion cues indicate movement (exceed certain level). Otherwise, pixel is marked as steady (black).

All arbitrary parameters, like window size and threshold levels depend on image resolution and noise intensity. These were optimized for experiments empirically.

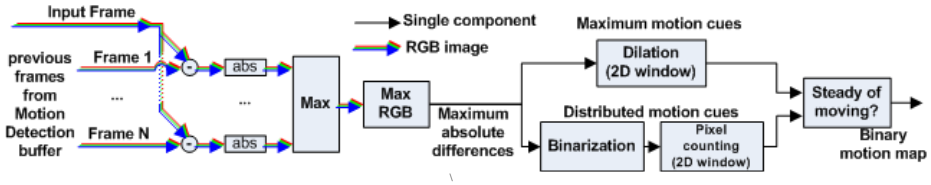

Fig. 3. Scheme of Motion Detector.

Fig. 4. Motion map (right) obtained for exemplary frame (left) (white pixels – moving, black pixels – steady).

3.2 Filtering

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 (Fig. 5a) with respect to previously filtered frame, stored in "Denoised" pipeline of the algorithm (Fig. 1). 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.

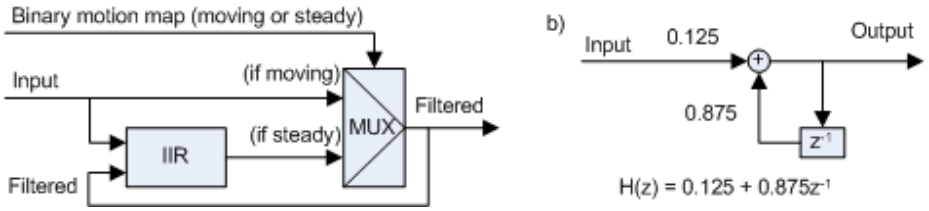

Fig. 5. Filtering scheme (a) and used low-pass IIR filter (b).

In our work we have exploited simple low-pass IIR (Infinite Impulse Response) filter of first-order (Fig. 5b). Low order filter was chosen to reduce computational complexity and to allow slight changes in the scene (e.g. day-time lighting).

3.3 Artifact removal

Noise removal scheme composed of presented Motion Detection and Filtering modules is simple and computationally efficient. Unfortunately, it may be a cause of artifacts resulting from hard-decisive classification of pixels as steady or moving.

Fig. 6 shows three trajectories of exemplary pixel: original value *(a)*, filtered value *(b)* and value after artifact removal *(c)*. At the beginning (segment I), the pixel is classified as steady. It varies due to noise, which is filtered (filtered *(b)* is the same as *(c)*). Then (II), pixel value starts to change significantly and is classified as moving. As a result of that, filtering phase is omitted: *(a)*, *(b)* and *(c)* are the same. Up to this moment, there are no artifacts.

In segment III, pixel is classified as steady, because its value changes very slowly. Filtered pixel trajectory changes even slower, resulting in discrepancy between trajectories (Fig. 6), which is lesser than threshold of motion detector. After a while, the discrepancy rise beyond threshold and pixel is instantaneously classified as moving in segment IV. Filtering switches off, and thus trajectories are updated to original, which causes another steady segment V. Rapid switching causes visual artifact in the output image.

To tackle that, we introduce an artificial removal step. Rapid changes of pixel classification (steady or moving) are predicted, with use of artifact prediction buffer (Fig. 2). If such rapid change is predicted, pixel value is linearly interpolated between original *(a)* and filtered *(b)* trajectories before the change occurs.

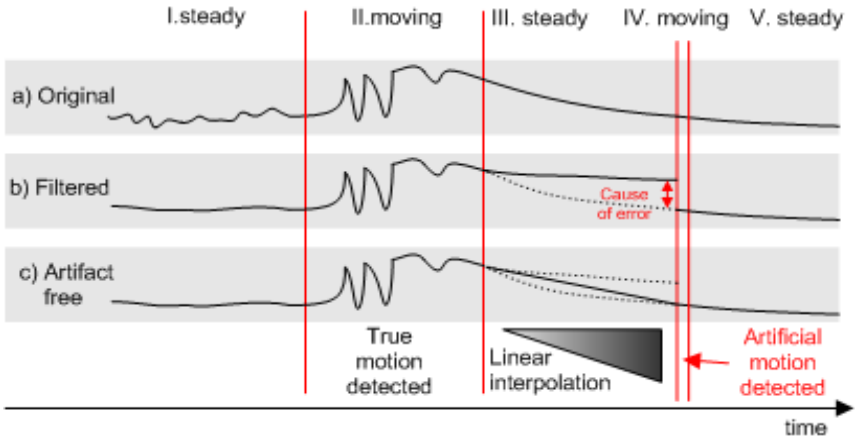

Fig. 6. Example of artifact removal on trajectories of exemplary pixel value.

4 Results

We have evaluated proposed approach to estimation of temporally consistent depth maps by use of noise removal experimentally. To assess influence of presented noise reduction technique on quality of depth data, depth maps have been estimated from

denoised and from original video. The tests have been performed on some of multi-view video sequences that are currently used in 3D standardization [10,11]. A state-of-the art depth estimation algorithm based on graph cuts [9] implemented in ISO/IEC MPEG Depth Estimation Reference Software have been used. For noise reduction, our presented simple technique was used.

Quality of depth maps was assessed indirectly, by assessment of quality of synthesized views (Fig. 7). These views have been synthesized with use of depth maps (generated form original and denoised video) and original sequences (even when depth maps were obtained from denoised version).

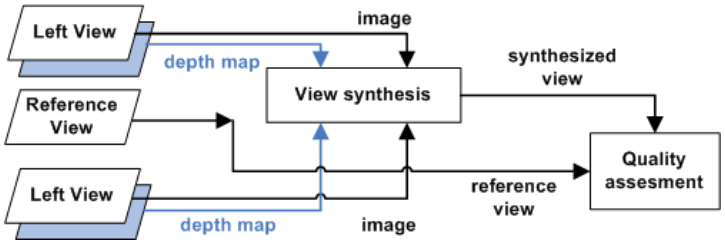

Fig. 7. Depth map quality assessment by assessment of quality of synthesized view.

Quality have been evaluated objectively (PSNR - Fig. 8 on vertical bars) and subjectively (15 subjects, MOS - Fig. 8) in comparison with original views. In case our study, MOS (Mean Opinion Score) is expressed by a 10-point continuous scale. Rating of the quality was in range from 1 ("very bad with annoying impairments/artifacts") to 10 ("excellent, artifacts are imperceptible").

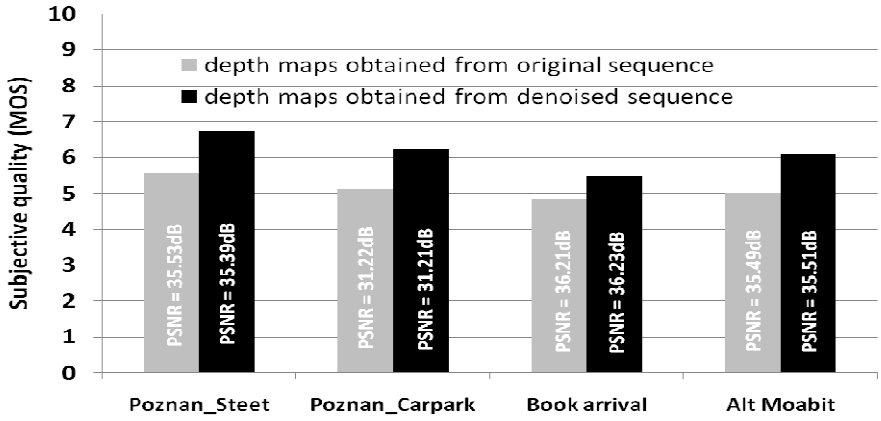

Fig. 8. Depth map quality assessment by assessment of quality of synthesized view.

Results presented in Fig. 9 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 9 show exemplary results attained with and without use of proposed approach. As can be noticed on Fig. 9a, 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. 9b,c) has not changed, temporal consistency expressed as difference between frames (Fig. 8d) 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.

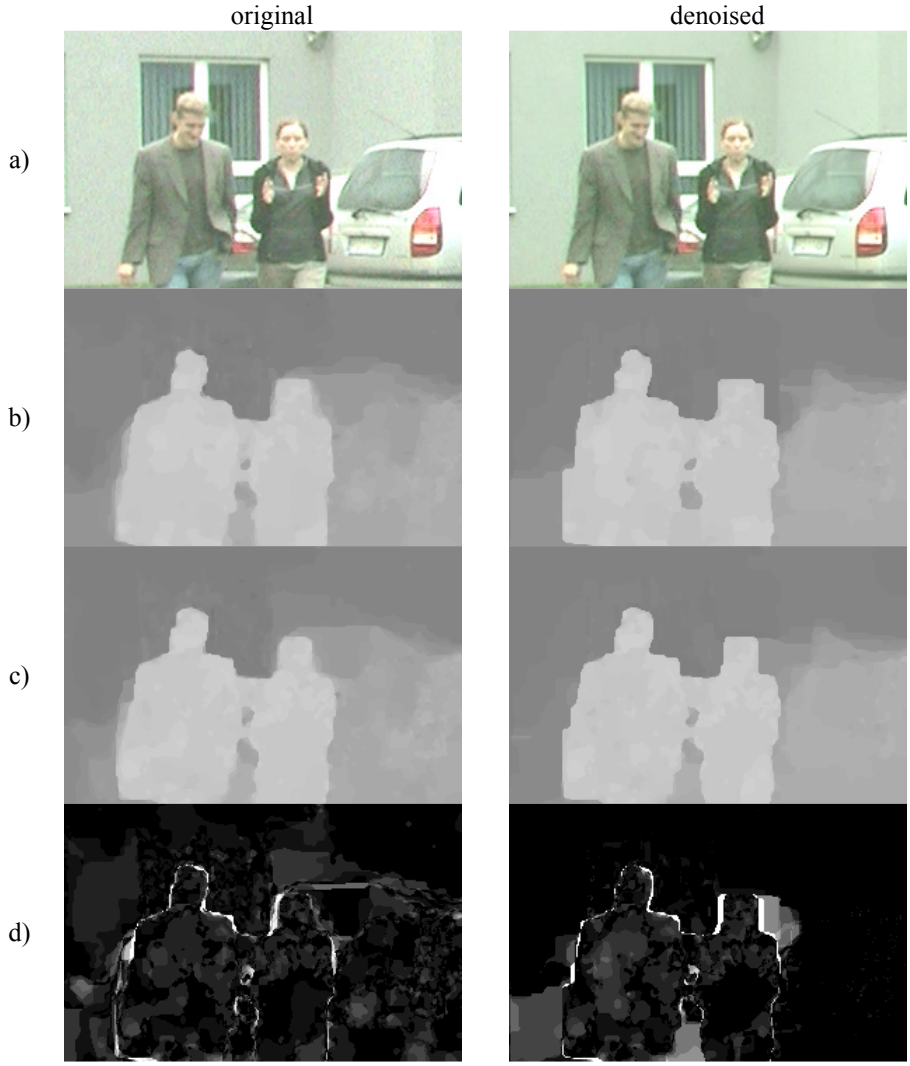

Fig. 9. Exemplary results of proposed technique: original (left) and denoised (right).
a) image b,c) depth maps for two consecutive frames d) difference between depth maps.

4. Conclusions

A novel approach to providing temporally consistent depth maps has been presented. As has been shown, use of devised technique significantly increases subjective quality of depth maps. Although objective quality is not altered, proposed approach can be successfully used for estimation of depth maps, where temporal consistency is desired, like in 3D television systems. An interesting new direction of work is to test proposed approach with use of different algorithms for depth estimation and more advances noise removal techniques. In this work, only an exemplary setup with simple noise reduction technique was used. It may be worth to extend presented denosing technique in future, to support moving objects instead of leaving them unchanged.

References

1. M. Domański, K. Klimaszewski, J. Konieczny, M. Kurc, A. Luczak, O. Stankiewicz, K. Wegner, „An experimental Free-view Television System", 1st International Conference on Image Processing & Communications, Bydgoszcz, Poland, Sept. 2009.
2. D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," Intern. Journal of Comp. Vision, vol. 47, pp. 7–42, 2002.
3. E. Scott Larsen et al., "Temporally Consistent Reconstruction from Multiple Video Streams Using Enhanced Belief Propagation", International Conference on Computer Vision 2007.
4. F.F. Pedro, P. H. Daniel, "Efficient Belief Propagation for Early Vision", International Journal of Computer Vision Vol. 70, No.1, Oct. 2006
5. H. Tao, H.S. Sawhney, R. Kumar, "Dynamic Depth Recovery from Multiple Synchronized Video Streams", International Conference on Image Processing 2003.
6. S. V. Vaseghi, "Advanced Digital Signal Processing and Noise Reduction (Third Edition)", John Wiley & Sons 2006, ISBN: 978-0-470-09495-2, 2006.
7. W.C. Van Etten, "Introduction to Random Signals and Noise", John Wiley & Sons 2006. ISBN 0-470-02411-9,2006.
8. R. Dugad, N. Ahuja, "Video denoising by combining Kalman and Wiener estimates" Proceedings of International Conference on Image Processing, 1999, pp 152-156.
9. O. Stankiewicz, K. Wegner, M. Wildeboer, "A soft–segmentation matching in Depth Estimation Reference Software (DERS) 5.0", ISO/IEC JTC1/SC29/WG11 MPEG/M17049, October 2009, Xian, China.
10. M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznan Multiview Video Test Sequences and Camera Parameters", ISO/IEC JTC1/SC29/WG11 MPEG/M17050, October 2009, Xian, China
11. 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.