This paper is published in the open archive of Mid Sweden University
DIVA http://miun.diva-portal.org
by permission of the publisher

Yun Li, Mårten Sjöström, Ulf Jennehag, and Roger Olsson, "Depth Image Post-processing Method by Diffusion," In Proceedings of SPIE - The International Society for Optical Engineering: Three-Dimensional Image Processing (3DIP) and Applications II, San Francisco, CA, Feb. 2013.
.
http://dx.doi.org/10.1117/12.2003183

© Copyright 2013 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic electronic or print reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

Depth Image Post-processing Method by Diffusion

Yun Li, Mårten Sjöström, Ulf Jennehag, and Roger Olsson

Dept. of Information Technology and Media
Mid Sweden University
SE-851 70 Sundsvall Sweden

ABSTRACT

Multi-view three-dimensional television relies on view synthesis to reduce the number of views being transmitted. Arbitrary views can be synthesized by utilizing corresponding depth images with textures. The depth images obtained from stereo pairs or range cameras may contain erroneous values, which entail artifacts in a rendered view. Post-processing of the data may then be utilized to enhance the depth image with the purpose to reach a better quality of synthesized views. We propose a Partial Differential Equation (PDE)-based interpolation method for a reconstruction of the smooth areas in depth images, while preserving significant edges. We modeled the depth image by adjusting thresholds for edge detection and a uniform sparse sampling factor followed by the second order PDE interpolation. The objective results show that a depth image processed by the proposed method can achieve a better quality of synthesized views than the original depth image. Visual inspection confirmed the results.

Keywords: Depth image, post-processing, view synthesis

1. INTRODUCTION

Depth based image rendering (DIBR) as a modern approach for three dimensional television (3DTV) rendering can generate multiple views at the receiver end with much less overhead than simulcast; this approach also enables backward compatibility for the existing 2D environment.[1] Intermediate virtual views are then rendered by using DIBR approach with the Multi-View video plus Depth (MVD) format data. The depth images of the MVD data can be assumed as piece-wise smooth gray level images that reflect the depth of the scene for their corresponding video texture images. Different post-processing methods have been devised to improve the depth image quality by using the piece-wise smooth assumption. A question that arises is whether synthesized views can be of better quality by eliminating insignificant edges and setting smoothness constraints on the depth image.

The paper by Scharstein et al.[2] provides taxonomy of methods based on stereo to disparity and compares their performance. Stereo matching algorithms, e.g. color segmentation, belief propagation,[3] and visibility constraints[4] can produce a coarse disparity map. Many different post-processing methods have been proposed over the years to improve a first-hand depth image obtained from stereo pairs or from range cameras. In the research work of Gangwal et al.,[5] depth images were down-sampled and then up-sampled using joint-bilateral filters to better align the object boundaries in the depth image and in its correlated texture. The method by Yang et al.[6] employs hierarchical joint-bilateral filters from a down-sampling depth image with matching confidence. Silvaa et al.[7] utilized Edge-adaptive joint trilateral filter and bilateral sharpening filter to combat block-based compression artifacts. Adaptive cross trilateral median filter[8]has also been introduced, it adapts to the local image structure and reduces the effect of mismatching.

In this paper, we propose a post-processing method that adds a smoothness constraint to depth images with aligned edges to the texture. The depth image is modeled by edges and uniform sparse sampling points and then reconstructed by solving Laplace equation with the least square error method. We investigate the view synthesis quality as a function of the identified edges and the uniform sparse sampling points on the corresponding depth image. The novelty of this paper is the application of PDE-based interpolation to depth image post-processing,

Further author information: (Send correspondence to, Mårten Sjöström)
E-mail: marten.sjostrom@miun.se, Telephone: +(46) 060 14-8836.

Three-Dimensional Image Processing (3DIP) and Applications 2013, edited by Atilla M. Baskurt, Robert Sitnik, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8650, 865003
© 2013 SPIE-IS&T · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2003183

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms

yet preserving the significant edges and geometrical relationship for a better quality of synthesized views. We define significance of the edges by the magnitude of the first order derivative of the depth image.

The sequel of the paper is organized as follows: Section 2 presents the scope and aim of this work. We describe our proposed reconstruction model in Section 3 and test set-up in Section 4. Results are showed in Section 5, and Section 6 concludes our work.

2. PROBLEM DESCRIPTION

Most depth images of a scene have distributions that consist of smooth areas and sharp transitions where a foreground object appears in front of background objects. Therefore, it is important to preserve these significant edges and assure smoothness in the areas between these edges. However, restrictions of depth image acquisition and existing post-processing techniques implies that inaccuracies will appear. The process of rendering an intermediate virtual view using the MVD format is further complicated by the disocclusion handling process, camera set-up and the precision of camera parameters etc. It is a great challenge to find appropriate post-processing to improve the quality of the depth image, so that the synthesized virtual view is closer to a true view at the corresponding position.

The aim of this work is to improve the quality of depth images for a better 3DTV experience. In this study, we assume that the original depth image in the MVD sequences has accurate edges, such that they coincide well with the associated texture edges. I.e. the problem of edge alignment is assumed to have been carried out by other post-processing algorithms, and so is out of the scope of this study. The goal is to investigate the quality of the reconstructed depth image and the synthesized view by using the proposed post-processing method.

3. PROPOSED METHOD

We model a depth image by a certain number of significant edges and a uniform sparse sampling. The significant edges are identified by applying a Canny edge filter with given thresholds. The uniform sparse sampling points are obtained by sub-sampling the original depth image by a constant factor. The improved depth image is reconstructed from pixels on both sides of the detected significant edges and the uniform sparse sampling points by applying a second order Partial Differentiation Equation-based (PDE-based) interpolation, namely a discrete version of the Laplace equation. The unknown of the equations are then solved by a least square error method.

The significant edges retain object borders and preserve important depth transitions, and the uniform sparse sampling points govern the smoothness versus the fidelity of the reconstruction, while the interpolation enforces the smoothness constraint for areas between the significant edges. See Figure 1 for the reconstruction and evaluation system.

The blocks *Canny detector*, *Edge extraction*, *Sparse sampling* and *Diffusion* are utilized in the proposed depth image post-processing. The details of each block are explained below:

Canny detector: The proposed method starts with edge detection. The canny edge detector[9] identifies the horizontal, vertical and diagonal edges, and applies two thresholds to find connected edges. In Figure 1, edges are detected by using Canny edge detector with thresholds, which are changed by a threshold multiplication factor T_e. Increasing the factor T_e decreases the detection sensitivity and will produce fewer edges of higher significance.

Edge extraction: Once the significant edges have been detected, the scheme needs to obtain pixels on both sides of the edges. For this purpose, an edge mask E_m is generated by morphological dilation

$$E_m = E_d \oplus S_3, \tag{1}$$

with a square structure element S_3 of size 3, where E_d is the edges detected in the previous step, and $\oplus$ is the morphological dilation. Pixels within the edge mask are selected, along with the border of the depth image for later diffusion.

Sparse sampling: The scheme also requires uniform sparse sub-sampling points of a sampling factor T_s from the original full depth image.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms

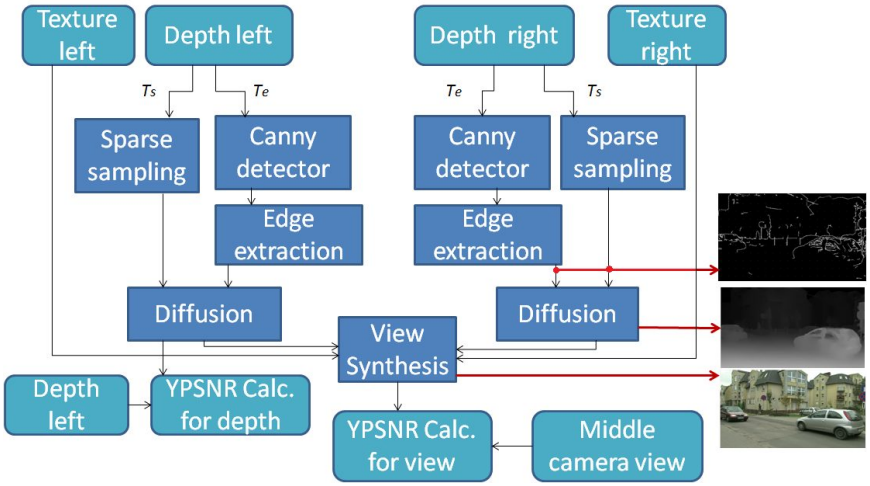

Figure 1. Evaluation system for depth image post-processing.

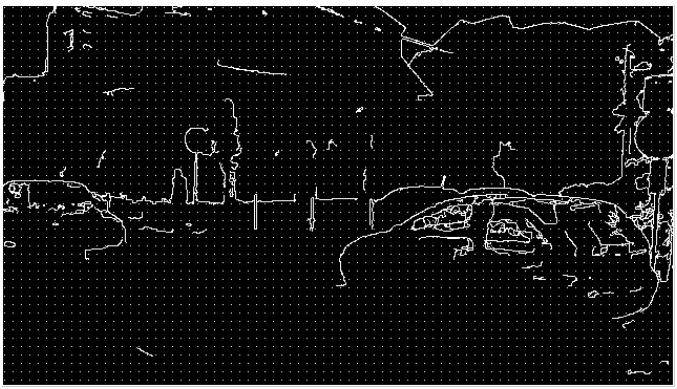

Figure 2. an example of the edge information and the uniform sparse sampling points required for diffusion.

Diffusion: The sparse sampling points and the extracted edge information are diffused to reconstruct the depth image. The extracted edge information includes pixels on the edges, pixels on both sides of the edges, and the border of the image. Figure 2 shows an example of the required information for diffusion.

The diffusion is carried out by applying the Laplace equation

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0, \tag{2}$$

on smooth areas of the depth image. In our implementation, the solution is approximated by

$$\begin{aligned} f(x,y) \quad = \quad & [f(x-1,y) + f(x+1,y) + \\ & f(x,y-1) + f(x,y+1)]/4. \end{aligned} \tag{3}$$

Eq. 3 implies that pixels of unknown are equal to the average of their horizontal and vertical four surrounding pixels. The equations are solved by the least square error method.

In this proposed method, the quality of the reconstructed depth image $\hat{f}(T_e, T_s)$ is varied by two parameters: the edge threshold multiplication factor T_e and the uniform sub-sampling factor T_s. The rational for such a modeling is that the extracted edges and the uniform sparse sub-sampling points are important information in depth images. The extracted edges contains essential depth values at the depth transitions with a certain

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms

significance, and these edges also correspond to objects boundaries in the corresponding video texture; the uniform sparse sub-sampling points ensure a consistency of depth between adjacent given views. In addition to the piece-wise smooth assumption of depth images, the consistency is necessary to be considered because, in the view synthesis process, the positions of virtual rendered pixels m_{vl} and m_{vr} from left and right cameras calculated by Eqs. (5) and (6) respectively should satisfy the constraint[10] in

$$m_{vl} = m_{vr}, \tag{4}$$

where

$$m_{vl} = \frac{z_l K_v R_v (K_l R_l)^{-1} m_l + (C_l - C_v)}{z_v}, \tag{5}$$

$$m_{vr} = \frac{z_r K_v R_v (K_r R_r)^{-1} m_r + (C_r - C_v)}{z_v}, \tag{6}$$

$$z = 1/(\frac{d}{255}(\frac{1}{z_{min}} - \frac{1}{z_{max}}) + \frac{1}{z_{max}}). \tag{7}$$

The depth $z_{(\cdot)}$ is computed from the pixels value d in the depth image, where $K_{(\cdot)}$ is the intrinsic parameter of the camera, $R_{(\cdot)}$ is the rotational matrix, and $C_{(\cdot)}$ the camera translation in Euclidean coordinates. The subscript v represents virtual view, l represents left camera, and r right camera. z_{min} and z_{max} is the minimum and the maximum depth of the scene, respectively. The disagreement of the relationship in Equation 4 will lead to artifacts on the synthesized view, in which double pixels of the same scene appear due to warping from the left and the right adjacent views, especially the corresponding texture contains edges on smooth areas of the depth image.

4. TEST ARRANGEMENT AND EVALUATION CRITERIA

The quality of the proposed post-processing method was evaluated by comparing a synthesized view rendered using the reconstructed depth image to the corresponding true view at the same camera position. We used different edge filter thresholds T_e and different sub-sampling factors T_s to investigate what values would give a rendered view closer to the original.

Two video sequences, Poznan Street and Poznan Hall2 from Poznan University of Technology,[11] were selected for the evaluation. Virtual views were synthesized by the MPEG View Synthesis Reference Software (VSRS)[12] version 3.5 at the centre position between two given camera positions (position 4 from camera position 3 and 5, and position 6 from camera position 5 and 7, respectively). The centered virtual views were rendered because those camera views are available. Furthermore the warping distances from the adjacent camera views are equally long, which make the center position the worst case to investigate. The quality was measured over 100 frames for each sequence.

We also compared the resulting improved depth images to the original depth images and the virtual views synthesized from the improved depth images to the true camera views. The objective evaluation was performed in Matlab. Figure 1 outlines the entire evaluation system. In addition, the rendered views were also visually inspected.

The Canny edge thresholds were changed by multiplying T_e with the default edge thresholds, which were detected by using Canny edge detector with default parameters in the Matlab. The default thresholds were [0.0125 0.0313] and [0.0063 0.0156] for Poznan Street and Poznan Hall, respectively. T_e was set to range from 0.25, 0.5, and 1 to 4; the uniform sparse sub-sampling factors T_s was assigned to 2, 4, 8, 16, 32 and none (no sampling). We considered YPSNR as a metric for our objective evaluation due to its simplicity and the high acuity of human visual system to the luminance component of an image.

$$\text{YPSNR} = 20\log_{10}(\frac{255}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(\text{MSE}_{Y_n})}}), \tag{8}$$

N is the number of frames, MSE_{Y_n} is the mean square error for the luminance of the n^{th} frame.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms

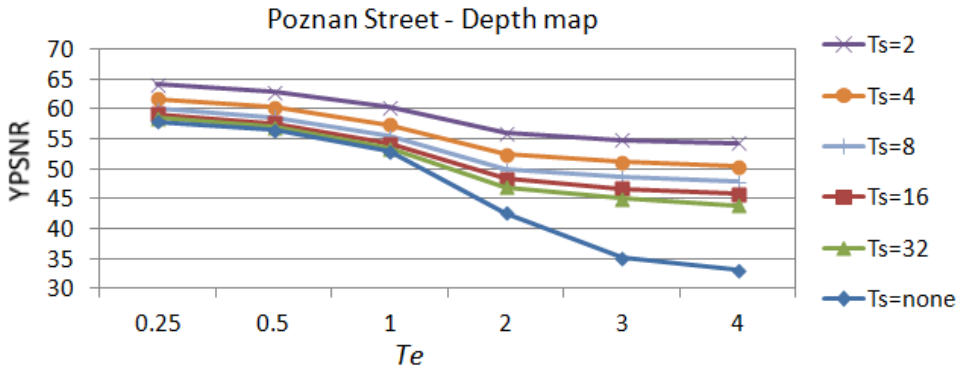

Figure 3. Objective quality for reconstructed depth image of Poznan Street.

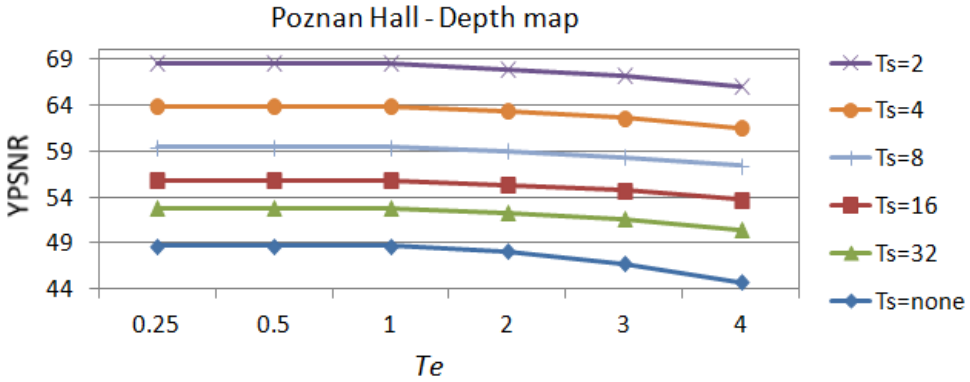

Figure 4. Objective quality for reconstructed depth image of Poznan Hall.

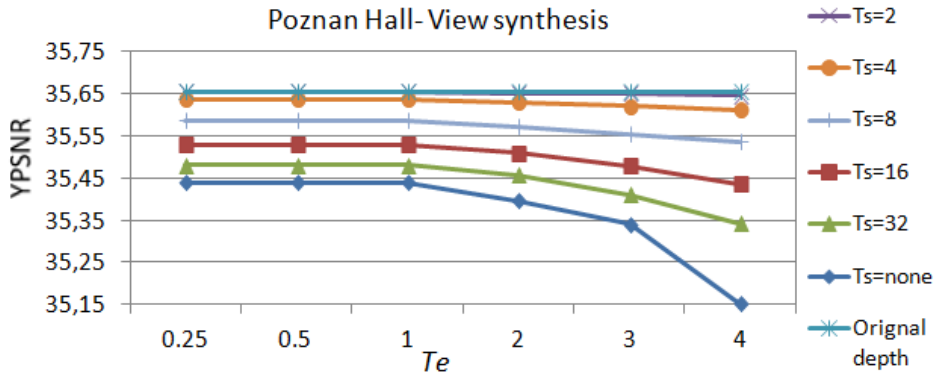

Figure 5. Objective quality for view synthesis of Poznan Hall.

5. RESULTS AND ANALYSIS

Figure 3 depicts the depth image reconstruction quality by the proposed method. YPSNR varies from 33 dB to 64 dB when changing T_e and T_s in the entire investigated range for Poznan Street. An increasing edge threshold and an increasing uniform sparse sampling factor both introduce more deviations from the original depth image. This is manifested in a reduction in PSNR for the reconstructed depth image relative to the original depth image). Similar results can be found in Figure 4 for Poznan Hall, which exhibited reconstruction quality ranged from 45 dB to 68 dB. Because the sequence Poznan Hall contains fewer edges and larger planar areas, the YPSNR is less sensitive to the edge thresholds and instead more affected by the uniform sparse sampling factor. For

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms

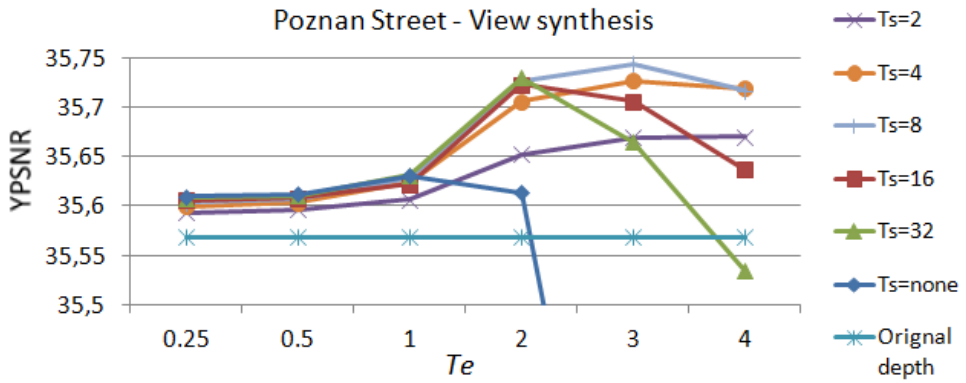

Figure 6. Objective quality for view synthesis of Poznan Street.

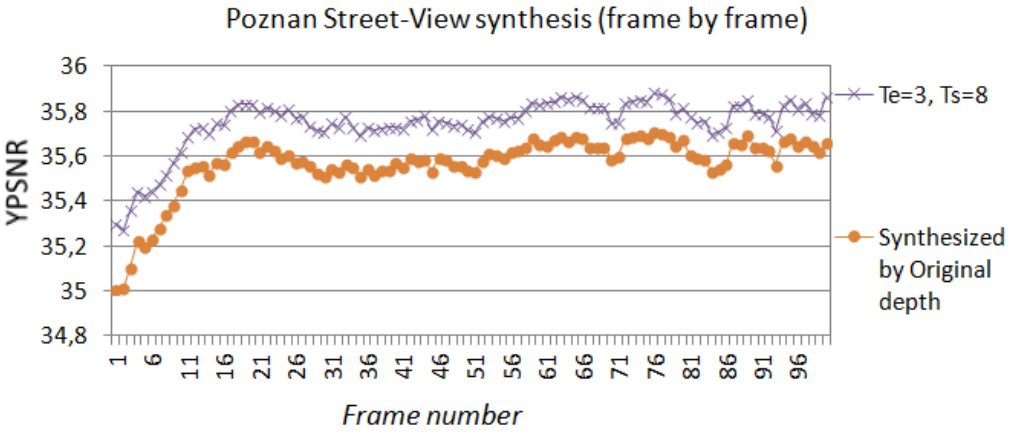

Figure 7. Objective quality for view synthesis of Poznan Street (frame by frame).

both sequences, more samples in smooth areas (small T_s) are required for a satisfactory quality of depth image reconstruction with fewer edges (large T_e).

Considering the synthesized views, the proposed post-processing method achieved an improvement in quality of 0.18dB for the Poznan Street sequence when the parameters values $T_e = 3$ and $T_s = 8$ were chosen. See Figure 6. Figure 7 further shows that the improvement is consistent frame by frame for $T_e = 3$ and $T_s = 8$. However, the proposed method did not demonstrate any improvements when applied to the Poznan Hall2 sequence. This may be due to the depth distribution in the scene of the sequence: it has a very simple structure characterized by clear edges and planar areas bounded by the edges. The proposed method cannot make the areas between the significant edges even smoother, and consequently no improved quality is registered in the evaluation. Our visual inspection also confirms the improvement in Poznan Street sequence. Figure 8 presents the details of the synthesized views.

6. CONCLUSION

In this paper, we proposed a depth image post-processing method by using PDE-based diffusion. The method models a depth image as significant edges and uniform sparse sampling points between these edges. This approach is founded on the fact that most depth images have a smooth distribution except at boundaries of different objects in a scene. An appropriate threshold for the Canny edge detector selects only significant edges. Insignificant edges are eliminated as the method applies a PDE-based interpolation for reconstructing the depth image. Thus,

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms

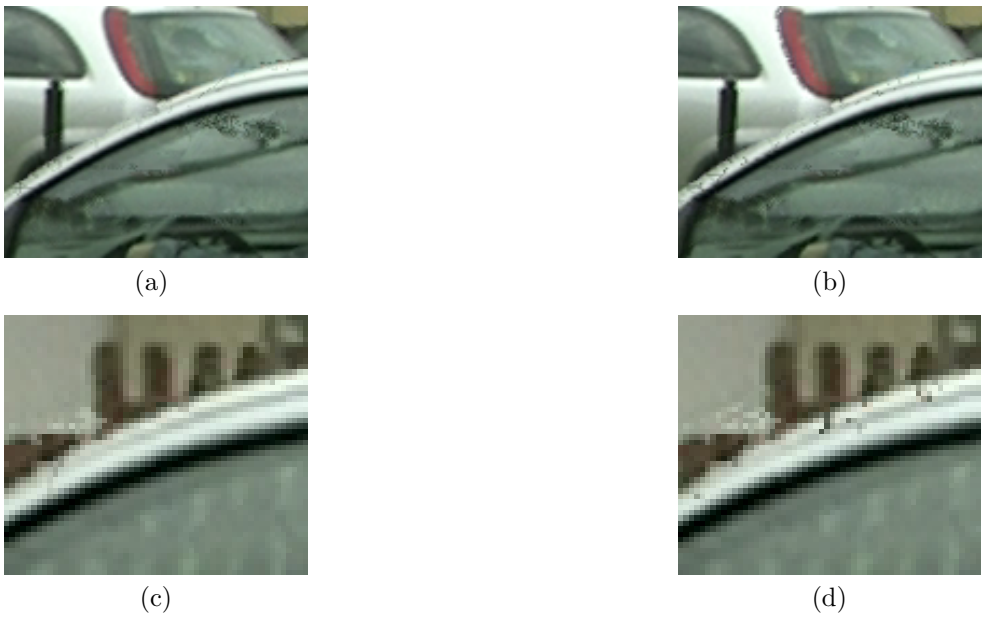

Figure 8. Details of the synthesized views for Poznan Street; (a) and (c) are from the 1^{st} and 2^{nd} frame synthesized using reconstructed depth image of $T_e = 3$ and $T_s = 8$; (b) and (d) are from the 1^{st} and 2^{nd} frame synthesized using original depth image.

this procedure increases the smoothness in the areas between the significant edges. The method preserves the geometrical relationship between views for a better quality view synthesis by keeping a set of sparse sampling points in the smooth areas.

The experiment results demonstrated that the depth image reconstruction quality varied with the parameter values, more deviations from the original depth image were introduced with an increasing edge threshold and an increasing uniform sparse sampling factor. The quality of synthesized views is improved by using the proposed post-processing method with appropriate parameter values on depth images with many edges of different significance. In cases when the depth image contains few edges and large smooth areas, little or no improvements will be observed. The improvement was confirmed by our visual inspection. A requirement for the proposed method is a good agreement of edges between depth image and its corresponding texture image.

ACKNOWLEDGMENTS

This work has been supported by grant 2009/0264 of the Knowledge Foundation, Sweden, by grant 00156702 of the EU European Regional Development Fund, Mellersta Norrland, Sweden, and by grant 00155148 of Länsstyrelsen Västernorrland, Sweden.

REFERENCES

[1] Olive, S., Peter, K., and Thomas, S., "3DTV Broadcasting," in [*3D videocommunication*], pp(23–36), John Wiley & Sons Ltd. (2005).

[2] Scharstein, D., Szeliski, R., and Zabih, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)* (1), 131–140 (2002).

[3] Klaus, A., Sormann, M., and Karner, K., "Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure," *18th International Conference on Pattern Recognition (ICPR'06)* , 15–18 (2006).

[4] Bleyer, M. and Gelaiitz, M., "A layered stereo algorithm using image segmentation and global visibility constraints," *International Conference on Image Processing (ICIP)* , 2997–3000 (2004).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms

[5] Gangwal, O. and Berretty, R., "Depth map post-processing for 3D-TV," in [*Consumer Electronics, 2009. ICCE'09. Digest of Technical Papers International Conference on*], 1–2, IEEE (2009).

[6] Yang, Q., Wang, L., Li, D., and Zhang, M., "Hierarchical Joint Bilateral Filtering for Depth Post-Processing," *2011 Sixth International Conference on Image and Graphics* , 129–134 (Aug. 2011).

[7] Silva, D. V. S. X. D., Fernando, W. A. C., Worrall, S., and Kondoz, A., "A Depth Map Post-Processing Technique for 3D-TV Systems based on Compression Artifact Analysis," *Signal Processing* , 1–30 (2011).

[8] Mueller, M., Zilly, F., and Kauff, P., "Adaptive cross-trilateral depth map filtering," *2010 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video* , 1–4 (June 2010).

[9] Canny, J., "A computational approach to edge detection.," *IEEE transactions on pattern analysis and machine intelligence* **8**, 679–98 (June 1986).

[10] Olive, S., Peter, K., and Thomas, S., "View Synthesis and Rendering Methods," in [*3D videocommunication*], pp(167–169), John Wiley & Sons Ltd. (2005).

[11] Domaski, M., Grajek, T., Klimaszewski, K., and Kurc, M., "Pozna Multiview Video Test Sequences and Camera Parameters," *ISO/IEC JTC1/SC29/WG11* (2009).

[12] ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631, "Report on experimental framework for 3D video coding," (Oct. 2010). Guangzhou, China.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 04/15/2013 Terms of Use: http://spiedl.org/terms