

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, Roger Olsson, and Tourancheau Sylvain "Subjective Evaluation of an Edge-based Depth Image Compression Scheme," In Proceedings of SPIE - The International Society for Optical Engineering: Stereoscopic Displays and Applications XXIV, San Francisco, CA, Feb. 2013.

<http://dx.doi.org/10.1117/12.2003053>

© 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.

# Subjective Evaluation of an Edge-based Depth Image Compression Scheme

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

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

## ABSTRACT

Multi-view three-dimensional television requires many views, which may be synthesized from two-dimensional images with accompanying pixel-wise depth information. This depth image, which typically consists of smooth areas and sharp transitions at object borders, must be consistent with the acquired scene in order for synthesized views to be of good quality. We have previously proposed a depth image coding scheme that preserves significant edges and encodes smooth areas between these. An objective evaluation considering the structural similarity (SSIM) index for synthesized views demonstrated an advantage to the proposed scheme over the High Efficiency Video Coding (HEVC) intra mode in certain cases. However, there were some discrepancies between the outcomes from the objective evaluation and from our visual inspection, which motivated this study of subjective tests. The test was conducted according to ITU-R BT.500-13 recommendation with Stimulus-comparison methods. The results from the subjective test showed that the proposed scheme performs slightly better than HEVC with statistical significance at majority of the tested bit rates for the given contents.

**Keywords:** Depth image compression, view synthesis, subjective test

## 1. INTRODUCTION

The Multi-view Video plus Depth (MVD) format is used to synthesize intermediate views for three dimensional televisions (3DTV). Depth images from the MVD data can be assumed as piece-wise smooth gray level images. The conventional video and image encoders (e.g. HEVC,<sup>1</sup> JPEG2000) will blur edges on the depth image, and this blurring introduces disturbing artifacts on virtual views synthesized from the decoded depth images.<sup>2</sup> We have previously devised an edge based depth image compression scheme<sup>3</sup> that can well preserve depth transitions in the depth image. The question is if this compression scheme can improve the quality of the synthesized view subjectively.

Objective video or image quality metrics (e.g. PSNR, SSIM) are less accurate for evaluating synthesized views than for decoded views.<sup>4</sup> We have performed an objective test considering SSIM for the edge based depth image compression scheme.<sup>3</sup> The testing results are inconsistent with the opinion from our visual inspection, which motivates this research work of subjective evaluation.

Various solutions have been proposed for depth image compression to preserve the properties of the depth image. Goto et al.<sup>5</sup> and Jager et al.<sup>6</sup> have proposed schemes that losslessly encode edge contours in a depth image and the coefficients of the polynomial, which is used to approximate the smooth areas bounded by the edge contours. A diffusion based scheme in the paper by Gautier et al.<sup>7</sup> applies bi-level image coding tool JBIG for coding of edge contours and employs differential pulse-code modulation for compressing pixel values on both sides of the contour and a sparse sampling. However, the above mentioned schemes have not been subjectively evaluated.

We have previously developed an edge-based scalable coding scheme for high quality depth image compression.<sup>3</sup> The scheme is based on lossless coding of edge contours, uniform sparse sub-samplings and a smooth inpainting. The scalability of the scheme is inherited from the coding structure, which assigns higher priorities

---

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

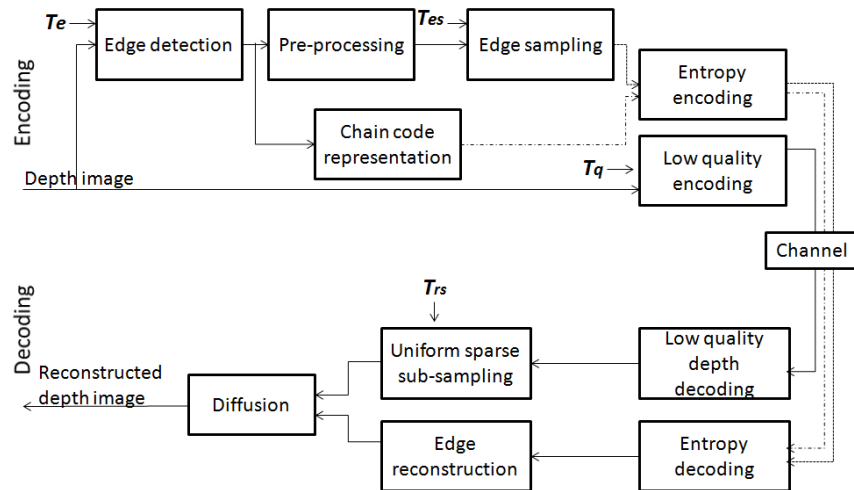


Figure 1. Coding system:<sup>3</sup> Encoding and Decoding subsystems.

to more significant edges. The significance of the edge is related to the magnitude of the first order derivatives of the depth image. A larger magnitude implies a higher significance.

In this research work, we conduct a subjective test for our proposed compression scheme. The test follows ITU-R BT.500-13 recommendation<sup>8</sup> using the stimulus-comparison method, which is selected because we assessed the quality of two synthesized images in relation with each other under the same encoded depth image bit-rate. The reference is the state of the art to be standard video codec, High Efficiency Video Coding (HEVC), which is very efficient in coding of video textures.

The overall aim is to improve depth image coding for a better perceived 3D experience. The work is limited in subjective evaluation of the proposed depth image compression scheme. The goal is to evaluate if better subjective quality on the synthesized views can be achieved by using the proposed scheme than the state of the art.

The remainder of the paper is organized as follows: Section 2 describes our previous image coding scheme, and the subjective evaluation method is illustrated in Section 3. Section 4 presents the results and analysis, and Section 5 concludes our work.

## 2. PREVIOUSLY PROPOSED CODING SCHEME

We briefly illustrate our previously proposed coding scheme here. A more detailed description of the scheme can be found in the research work of a scalable coding approach for high quality depth image compression.<sup>3</sup> The coding system showed in Figure 1 includes the *Encoding* and the *Decoding* subsystems.

### 2.1 Encoding

The scheme encodes edge contours, sub-sampling pixels on both sides of the contours and a low quality depth image. Each block of the encoding is illustrated as follows:

*Edge detection* detects edge contours from a source depth image by using Canny edge detector with a threshold multiplication factor  $T_e$ . A larger  $T_e$  decreases the detection sensitivity and results in fewer edges with higher significance.

*Chain code representation*: these edge contours are later represented by Freeman chain code. The representation includes starting location  $(X, Y)$ , length  $L$  and directions of the edge contours.

*Pre-processing* is applied on pixels around the edge contours to render more coherent depth values on each side of the contours. For this purpose, an edge mask is created by morphological dilation. The pixels within this mask are then removed, and their values are assumed to be equal to the values of their adjacent pixels, but, without crossing the edge contours. The unknowns are obtained by solving equations with a least square method.

*Edge sampling*: the pixels around these edge contours are sub-sampled uniformly by a constant factor  $T_{es}$ .

*Entropy encoding*: the values of the sub-sampled pixels and the directions of the edge contours are then differentiated and arithmetic encoded; the starting locations and the length of the edge contours are also arithmetic encoded.

*low quality encoding*: a low quality compressed data of the depth image is required to be transmitted to the decoder side, where uniform sparse samplings can be obtained from this low quality decompressed depth image.

## 2.2 Decoding

A reconstructed depth image is diffused from uniformly sub-sampling points, interpolated pixels around the edge contours and an image border in the decoding process. Each block of this process is described below:

*Low quality depth decoding*: the low quality compressed depth image is decoded.

*Uniform sparse sub-sampling*: the uniform sub-sampling points by using a factor of  $T_{rs}$  from the decoded low quality depth image and the border of this image are extracted for later diffusion.

*Entropy decoding*: edge information including the edge contours and the pixels around them is decoded.

*Edge reconstruction*: pixels around the edge contours are linearly interpolated. Pixels on the contour are filled by the interpolated pixels from right side of the contour.

*Diffusion*: the reconstructed depth image is diffused by solving the Laplace equation with a least square method from the extracted uniform sub-samplings, the border of the image, and the reconstructed edge information.

## 2.3 Scalability

The scalability is in relation to the significance of the edge contours in a depth image. Those edge contours that are detected with higher thresholds are more important than those that can be detected with lower thresholds; the less significant edge information should be discarded if the channel is subjected to a bandwidth constraint.

# 3. SUBJECTIVE EVALUATION

The subjective evaluation was conducted by following the ITU-R BT.500-13 recommendation.<sup>8</sup>

## 3.1 Test procedure

ITU-R BT.500-13 recommendation provides methodologies for image quality assessment. The recommendation consists of the Double-Stimulus Impairment Scale (DSIS) method, the Double-Stimulus Continuous Quality-Scale (DSCQS) method, and the alternative assessment methods. The DSIS and the DSCQS methods are used for assessing the quality of pictures in relation to unimpaired source pictures. In our subjective test, the quality of two impaired pictures in relation to each other was under consideration. Therefore, the Stimulus-comparison methods of the alternative assessment methods were utilized. The type of stimulus-comparison methods chosen was the Adjectival categorical judgment methods.

In the test, two impaired images in a pair from a source were presented to observers. They were asked to compare two images in relation to each other in the pair and enter scores in a paper scoring sheet. The scores were expressed as in Table 1. Figure 2 illustrates the presentation structure for test materials. Each image in the pair was presented on a display screen for 4 seconds. Between the presentations of two images in the pair, a mid-gray image was displayed for 3 seconds. Number of repetition for presenting each pair was 3, and voting took 6 seconds while showing a mid-gray image on the screen.

Table 1. Comparison scale

-3	Much worse
-2	Worse
-1	Slightly worse
0	The same
1	Slightly better
2	Better
3	Much better

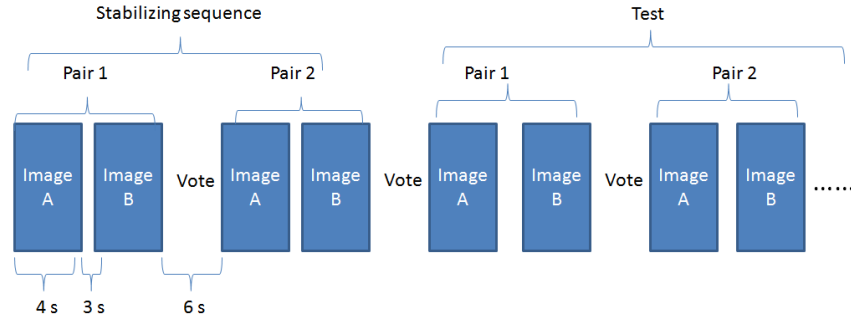


Figure 2. Presentation structure with durations of the presentation.

### 3.2 Apparatus and environment

A 23 inch full HD display Alienware2310 for two-dimensional viewing was utilized for the test. The display has a native resolution of  $1920 \times 1080$  pixels. The testing environment follows the ITU-R BT. 500, e.g. viewing distance: 3 times of display height, peak luminance of the screen:  $200 \text{ cd/m}^2$ , ratio of luminance of background behind picture monitor to peak luminance of picture: 0.15, chromaticity of background:  $D_{65}$ , room illumination: low.

### 3.3 Test material and error conditions

We chose MVD images from the MPEG video sequences<sup>9</sup> as test sources (SRC), i.e. SRC1: Poznan Street<sup>10</sup> (1<sup>st</sup> frame) and SRC2: Lovebird1 (239<sup>th</sup> frame). Virtual images were synthesized between the given MVD images using the MPEG View Synthesis Reference Software (VSRS)<sup>11</sup> version 3.5.

The rationale of choosing these two images is because the first frame from Poznan Street contains a complex scene, which is characterized by graduate transitions of the depth from background to foreground and with slightly larger depth discontinuities for the front most objects. The 239<sup>th</sup> frame of Lovebird1 contains large depth discontinuities at the boundaries of the front objects and a relatively smooth background. In addition, edges from these depth images coincide well with edges from their corresponding textures.

Images at virtual camera 4 were synthesized from camera 3 and 5 for Poznan Street and virtual camera 6 from 4 and 8 for Lovebird1 by VSRS. These virtual views were synthesized from original textures and decoded depth images. The depth images were compressed by using HEVC version 6.5 with quantization parameter (QP) 25, 27, 29, 31, 33 and 35, and by using our scheme with corresponding bit rates. Accordingly, the hypothetical reference circuits (HRC) were formed by processes of the depth image compression and the view synthesis, i.e. HRC1: "the proposed scheme+VSRS" and HRC2: "HEVC+VSRS".

Two synthesized images from the same test sequence were paired such that they were synthesized from the coded depth images of similar bit rates, and that they were produced from the proposed scheme and the reference encoding scheme (HEVC). Therefore, 12 pairs were resulted from the pairing process. In addition, the synthesized images were cropped for Poznan Street and padded for Lovebird1 to the native resolution of the display. The padding was in mid-gray color.

### 3.4 Test subjects, training and randomization

In this test screening, observers who have visual acuity below 0.8 or color blindness were unqualified for the subjective test. Visual acuity and color vision were examined by using the Snellen chart and the Ishihara chart respectively.

After the screening of 21 subjects, 18 naive subjects were qualified for the test. They were aged from 20 to 29, and 7 of them were female. The unqualified subjects also participated in the test, but results obtained from them were removed. All of the subjects were given written instructions before the test session, and these instructions were also explained by a test operator so that they fully understood how to proceed with the test. The graphical user interface and the examples of processed images were shown in a training section. One observer at a time was required to sit in front of the display for conducting the test.

The subjective test consisted of two test sessions, and the pairs from the same test source were presented in the same session. Two images in each pair were presented in a random order. Prior to the test, two stabilizing pairs were presented, and scores given to these pairs were excluded.

### 3.5 Analysis

Mean opinion scores and deviations with 95 % confidence interval are of our interest. The mean opinion score is calculated by

$$\bar{u}_{jk} = \frac{1}{N} \sum_{i=1}^N u_{ijk}, \quad (1)$$

where  $u_{ijk}$  is the score from assessor  $i$  under test condition  $j$  for image  $k$ , and  $N$  is the number of assessors. The associated confidence interval with the mean score is given by

$$[\bar{u}_{jk} - \delta_{jk}, \bar{u}_{jk} + \delta_{jk}], \quad (2)$$

where

$$\delta_{jk} = 1.96 \frac{S_{jk}}{\sqrt{N}}, \quad (3)$$

and the standard deviation is computed by

$$S_{jk} = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_{jk} - u_{ijk})^2}{N-1}}. \quad (4)$$

Outliers are found based on a criterion that if, for more than half of the tested pairs from one test source, the voting scores of a test subject deviate above twice the standard deviation, i.e.

$$|u_{ijk} - \bar{u}_{jk}| > 2S_{jk}, \quad (5)$$

the results from this individual are removed.

None of the 18 observers were outlier judged by the criterion. The scores obtained from them were taken into analysis. The analyzed results are plotted with the mean and the standard deviation of 95 % confidence interval. Results from the objective evaluation considering SSIM are also presented.

## 4. RESULTS AND ANALYSIS

Results for objective quality illustrated in Figure 3 and 5 are taken from our previous work.<sup>3</sup> The subjective test results are showed in Figure 4 and 6 for Poznan Street and Lovebird1 respectively.

The subjective tests demonstrated that the proposed scheme performs slightly better than HEVC for Poznan Street at all tested bit rates with statistical significance, except at the bit rate corresponding to HEVC QP 35, which is at 0.027 bit per pixel (bpp) (See Figure 4). For images from Lovebird1, the proposed scheme performs slightly better than HEVC at all tested bitrates, (See Figure 6). The results also indicated that SSIM

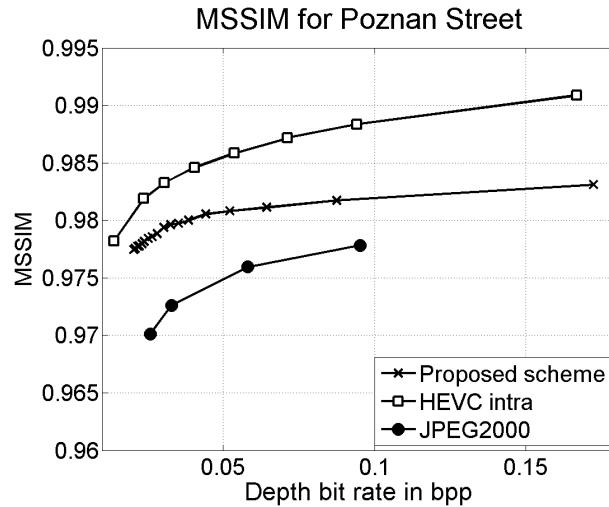


Figure 3. Objective quality for the synthesized views of Poznan Street.<sup>3</sup>

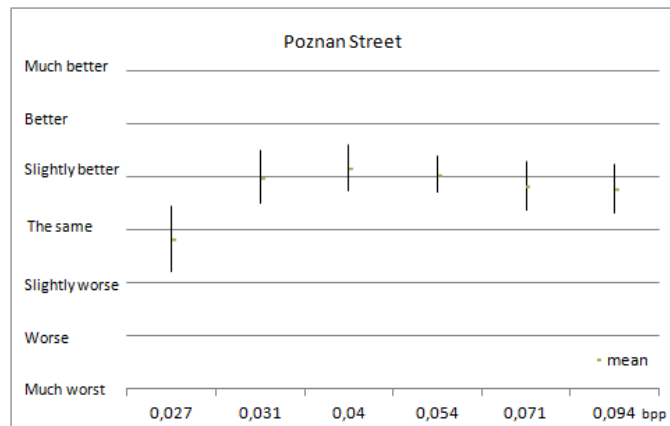


Figure 4. Subjective evaluation results for the synthesized views of Poznan Street (the proposed scheme versus the HEVC).

is inaccurate with respect to the perceptual quality on the synthesized view, (See Figure 3 and 5). Parts of the synthesized views are presented in Figure 7 and 8 for Poznan Street and Lovebird1 respectively. The quality of edges on the texture corresponding to the significant depth transitions was improved for the proposed scheme, and the scattering artifacts introduced by using the HEVC were reduced significantly.

The above experimental results show that the artifacts on the synthesized views introduced by the HEVC compressed depth images are more disturbing than by the proposed scheme. We interpret the results that preserved edges in a synthesized view is an important characteristic for quality of experience, as the proposed compression scheme preserves significant depth continuities in a depth image; it seems that distortion at edges in a synthesized view draws the attention of observers. However, HEVC is a rate-distortion optimized compression scheme, which guarantees the fidelity of the entire compressed depth image. Therefore, it introduces less distortion on the overall synthesized view than the proposed scheme at the very low bit rate (i.e. at 0.027 bpp) showed in Figure 4.

## 5. CONCLUSION

We have previously proposed a depth image coding scheme that preserves significant edges and encodes smooth areas between these. An objective evaluation considering the SSIM for synthesized views demonstrated an advantage to the proposed scheme over the HEVC intra mode in certain cases. However, some discrepancies

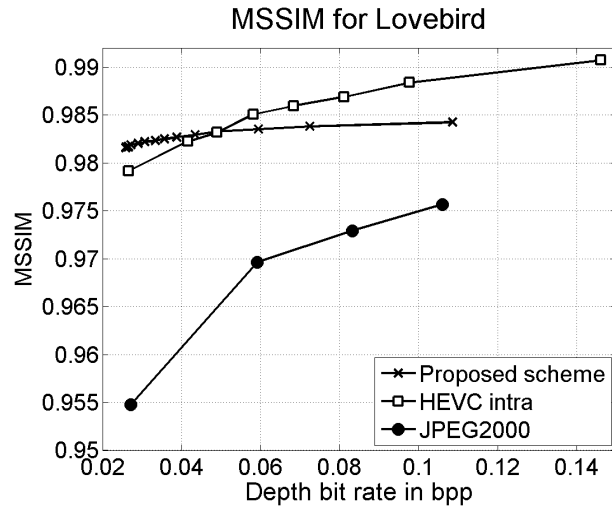


Figure 5. Objective quality for the synthesized views of Lovebird1.<sup>3</sup>

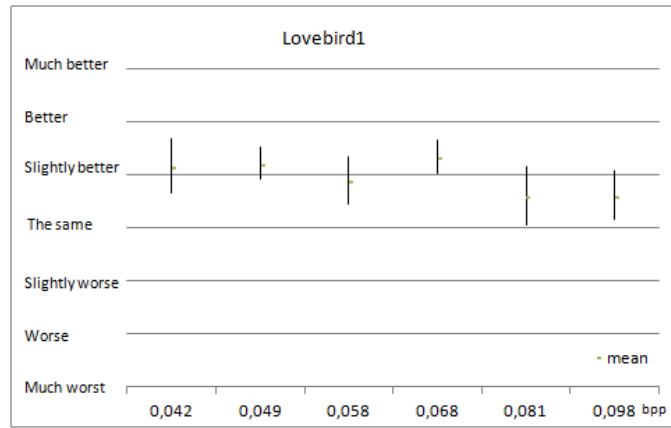


Figure 6. Subjective evaluation results for the synthesized views of Lovebird1 (the proposed scheme versus the HEVC).

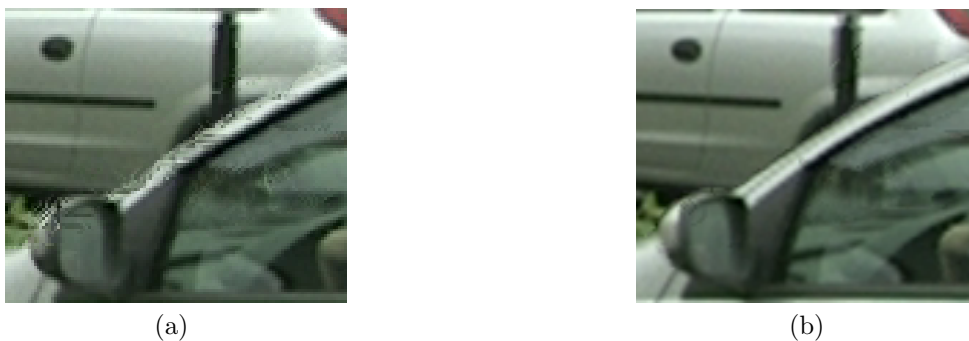


Figure 7. Parts of the synthesized views from Poznan Street by using (a) the HEVC and (b) the proposed method at the bit rate 0.027 bpp, respectively.



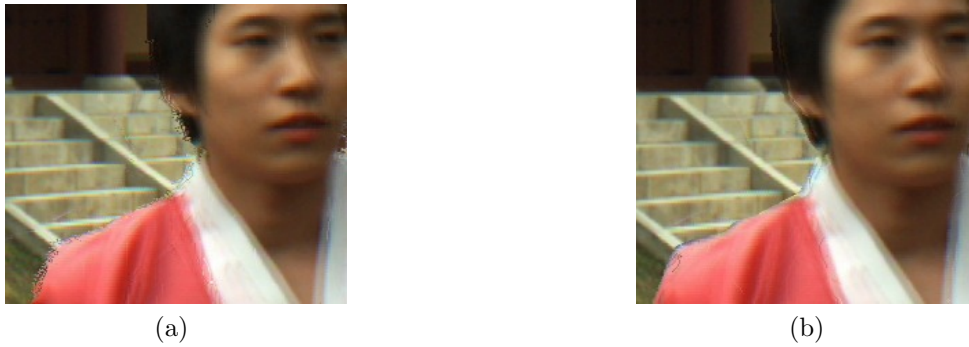


Figure 8. Parts of the synthesized views from Lovebird1 by using (a) the HEVC and (b) the proposed method at the bit rate 0.042 bpp, respectively.

exist between the objective evaluation and our visual inspection. Therefore, a subjective test was conducted according to ITU-R BT.500-13 recommendation with Stimulus-comparison methods. The synthesized images of the test sequences were paired such that they used encoded depth images of similar bit rates produced from the proposed scheme and HEVC respectively. The investigation demonstrates that the proposed compression scheme for depth images is preferred over using HEVC Intra coding, when a synthesized intermediate image is observed by test viewers. The research work reveals that preserving significant depth continuities in a depth image can result in an overall better quality of experience for synthesized views. It also illustrates that the objective metric SSIM fails to predict the importance of preserved edges for perceived quality in a synthesized view.

#### 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] JCT-VC, “WD1: Working Draft 1 of High-Efficiency Video Coding,” *JCTVC-C403* (Oct. 2010). Guangzhou, China.
- [2] Liu, S., Lai, P., Tian, D., Gomila, C., and Chen, C., “Joint trilateral filtering for depth map compression,” *Image Processing* **7744**, 77440F1–10 (2010).
- [3] Li, Y., Sjöström, M., Jennehag, U., and Olsson, R., “A scalable coding approach for high quality depth image compression,” in *Proc. 3DTV CON., Oct.* (Oct. 2012). Zurich, Switzerland.
- [4] Hanhart, P. and Ebrahimi, T., “Quality assessment of a stereo pair formed from decoded and synthesized views using objective metrics,” in *Proc. 3DTV CON.* (Oct. 2012). Zurich, Switzerland.
- [5] Goto, S., “Region-interior painting in Contour Based Depth map Coding System,” *2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, 1–5 (Dec. 2011).
- [6] Jager, F., “Contour-based segmentation and coding for depth map compression,” *2011 Visual Communications and Image Processing (VCIP)*, 1–4 (Nov. 2011).
- [7] Gautier, J., Meur, O., and Guillemot, C., “Efficient Depth Map Compression based on Lossless Edge Coding and Diffusion,” 81–84 (2012).
- [8] Recommendation ITU-R BT.500-13, “Methodology for the subjective assessment of the quality of television pictures,” (Jan. 2012).
- [9] ISO/IEC JTC1/SC29/WG11, “Call for Proposals on 3D Video Coding Technology,” (2011).
- [10] Domaski, M., Grajek, T., Klimaszewski, K., and Kurc, M., “Pozna Multiview Video Test Sequences and Camera Parameters,” *ISO/IEC JTC1/SC29/WG11* (2009).
- [11] ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631, “Report on experimental framework for 3D video coding,” (Oct. 2010). Guangzhou, China.