

This material is published in the open archive of Mid Sweden University

DIVA <http://miun.diva-portal.org>

to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Olsson, R.; Sjöström, M., "A Depth Dependent Quality Metric for Evaluation of Coded Integral Imaging Based 3D-Images," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2007* , pp. 1 - 4, 7-9 May 2007

<http://dx.doi.org/10.1109/3DTV.2007.4379397>

© 2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

A DEPTH DEPENDENT QUALITY METRIC FOR EVALUATION OF CODED INTEGRAL IMAGING BASED 3D-IMAGES

Roger Olsson and Mårten Sjöström

Dep. of Information Technology and Media, Mid Sweden Univ., SE-851 70 Sundsvall, Sweden

ABSTRACT

The two-dimensional quality metric Peak-Signal-To-Noise-Ratio (PSNR) is often used to evaluate the quality of coding schemes for integral imaging (II) based 3D-images. The PSNR may be applied to the full II resulting in single accumulate quality metric covering all possible views. Alternatively, it may be applied to each view results in a metric depending on viewing angle. However, both of these approaches fail to capture a coding scheme's distribution of artifacts at different depths within the 3D-image. In this paper we propose a metric that determines the 3D-image quality at different depths. First we introduce this 1D measure, and the operations that it is based on, followed by the experimental setup used to evaluate it. Finally, the metric is evaluated on a set of 3D-images; each coded using four different coding schemes and compared with visual inspection of the introduced coding distortion. The results indicate a good correlation with the coding artifacts and their distribution over different depths.

1. INTRODUCTION

The research into 3D-images and 3D-video is becoming more and more extensive. 3D-images captured by integral imaging (II) cameras are like other light field variants (e.g. multi-view) a data representation that is demanding in the context of storage and transmission. Thus, coding is next to mandatory and lossy coding is preferred to achieve data rates feasible for present transmission channels. To achieve an optimal trade-off between rate and distortion, a relevant quality metric is a vital tool in the coding process.

A commonly used metric when evaluating image and video coding schemes is the Peak-Signal-To-Noise-Ratio (PSNR). This also holds for 3D-images and -video. For example, when coding II-based 3D-images the PSNR has been applied to the full 3D-image directly as well as to single elementary images (EI), i.e. the subset of the pixels that are beneath a single lenslet [1, 2]. There are both pros and cons with metrics that result in a single accumulated value for a 3D-image. The advantage is that the quality of all possible views contained in the 3D-image is aggregated into a single easily comparable

value. However, a single value fails to capture how a coding scheme explicitly affects the 3D-image from specific viewing angles or at different scene depths. The quality metric proposed by Forman et. al [3] address this problem by extending the single valued PSNR into a PSNR-based quality matrix; each element in the matrix representing the quality of the 3D-image perceived by a virtual viewer from a specific viewing angle at an infinite viewing distance from the display.

Recently we have investigated a set of coding schemes that introduce codings artifacts that are scattered heterogeneously within the depth range of the 3D-image [4]. The depth varying distortion is a coding property that can neither be explicitly captured by a single global quality value nor by an angle dependent quality metric. In this paper we propose a quality metric that evaluates the quality of the 3D-image at different depths. The result, which is a 1D quality measure, provide a tool for investigating how a certain coding scheme affect the 3D-image in the depth dimension.

2. PROPOSED DEPTH DEPENDENT QUALITY METRIC

This section first gives a synoptic description of the proposed quality metric followed by three subsections giving a detailed description of the constituting parts.

Figure 1 illustrates the different steps that are performed when calculating the metric. As a full-reference metric it requires access to the coded and the original 3D-image. From these two 3D-images:

1. a set of 2D-image pairs are synthesized at different depth layers using image based rendering (IBR),
2. pixels that can be discerned to belong to each depth layer are identified and
3. the image pairs are masked and the coding artifacts at each depth layer is evaluated using a 2D quality metric.

Combined these steps result in a quality metric that is able to measures a coding scheme's depth distribution of artifacts. As can be noted from Figure 1 there are three operations that are the main parts of the proposed metric:

- Rendering

This work is supported by the Swedish Graduate School of Telecommunications and by the EU Objective 1 - programme Södra Skogslän region.

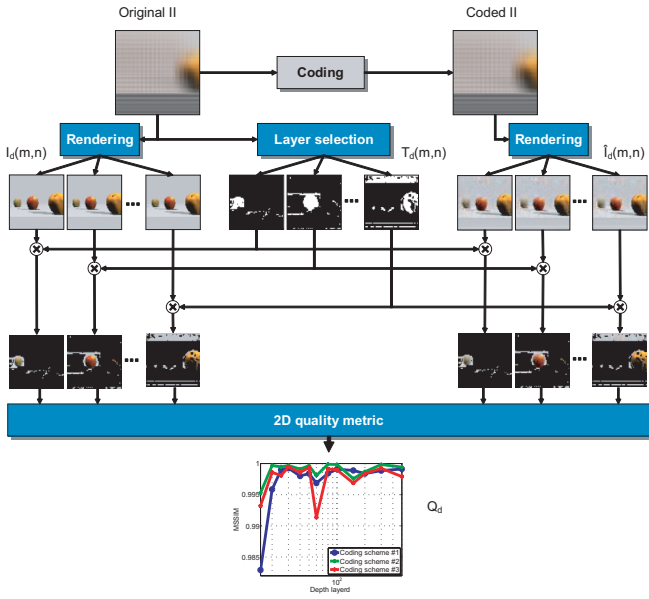


Fig. 1. Block diagram for the proposed depth enabled quality metric.

- Layer selection
- 2D-quality metric

2.1. Rendering

The image pair of each depth layer is synthesized using a plane-model IBR, i.e. novel virtual images are rendered using the EIs captured by the II-cameras lenslets. This section will elaborate on how we use IBR to synthesize a single depth layer from an II-based 3D-image.

To utilize the information captured by all lenslets in the II-camera, the field-of-view α of the virtual camera is set equal to the lenslets' field-of-view. This is achieved by positioning the virtual camera C_{IBR} on the optical axis at xyz-position $[0, 0, \frac{-f}{\Delta}]$ behind the image plane of the II-camera (at $z = 0$) according to Figure 2. The positioning of a depth layer d (parallel to the xy-plane) is normalized with respect to the II-camera such that it is a multiple of the lenslet focal length f .

When rendering the virtual image pair it is vital to preserve the coding artifacts located at that specific depth layer (from this virtual camera position) for the subsequent quality measurement. Therefore when calculating each pixel, only one lenslet's EI pixel should contribute to the color. Integrating over a set of pixels should be avoided since this has a low-pass filtering effect and would smear the coding artifacts. In Figure 2 this implies that the color of the image pixel $X_d(m, n)$ (marked red) is taken to be the color of a pixel in the EI covered by $C_{0,1}$. Which EI-pixel to use is derived by projecting the virtual image pixel's position onto the xy-plane

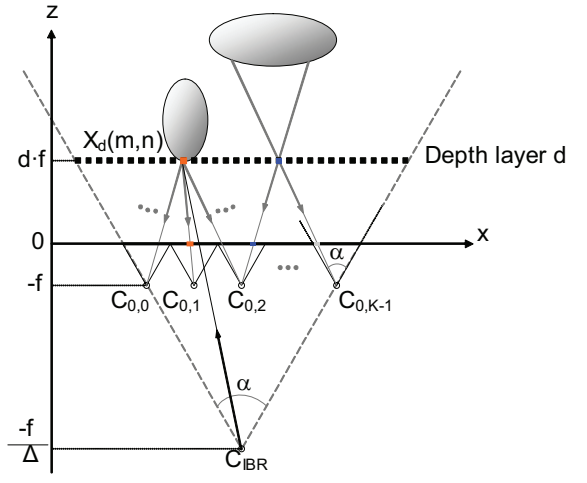


Fig. 2. Geometry of the virtual camera placement relative to the II-camera's lenslets. To simplify the figure only the xz-plane with the first row of lenslets ($C'_{0,k}$) is shown.

using the virtual camera's projection matrix. Linear interpolation is adopted to allow for non-integer EI-pixel positions.

2.2. Layer selection

Not all of the depth layer's rendered pixels correspond to objects that are intersected by the plane of that particular depth layer. A selection must be performed for the depth dependent metric to distinguish each layer. Pixels that correspond to objects outside the depth layer must be discarded prior to applying the 2D-quality metric.

To select which pixels to include we derive a depth map from the uncoded original 3D-image using the focus measure proposed by Takahashi and Naemura [5]. They conclude that synthesizing and combining a base image $B_d(m, n)$ with a reference image $R_d(m, n)$ – each derived by integrating over different sets of EIs – allows for calculating a measure of how likely it is for a pixel $X_d(m, n)$ to belong to depth layer d . Image pixels that correspond to a diffuse reflecting object located at the depth layer (e.g. the red pixel in Figure 2) will have similar EI-pixel contributions. Other image pixels (e.g. the blue pixel in Figure 2) will be an average of objects *outside* the depth layer and the EI-pixel contributions will vary. Evaluating the focus measure at different depths allows for a pixel-resolution depth map to be constructed by:

$$D(m, n) = \arg \min_d (|B_d(m, n) - R_d(m, n)| * h), \quad (1)$$

where h is a filter kernel that is used in a convolution step (operator $*$) to enhance the result, as described in [5].

Contrary to the rendering stage, this layer selection process benefits from including a *larger* set of EIs when rendering the base and reference images. Averaging over a larger

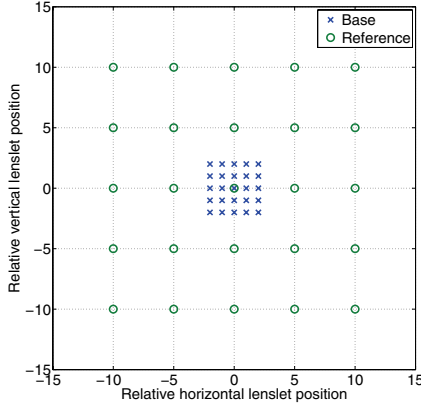


Fig. 3. Relative lenslet positions that are used to derive a specific image pixel for the base- and reference virtual image respectively.

set increases the reliability of the focus measure and consequently the accuracy of the depth map. Compared to the original focus measure definition we therefore increase the number of EIs that contribute to each image pixel from 2 to 25. Figure 3 shows the EIs patterns that are used to select which EI pixels (relative to the one found using intersection as described previously) that contribute to the color of the base and reference image. Based on the depth map D , a mask T_d (corresponding to a specific depth layer) is derived as:

$$T_d(m, n) = \begin{cases} 1 & \text{if } D(m, n) = d \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

2.3. 2D-quality metric

Any type of 2D image quality metrics could be used to measure the coding artifacts, e.g PSNR. However, by choosing the Mean Structural Similarity index (MSSIM) [6], its sliding-window principle allows for capturing inter-pixel effects that the pixel-to-pixel difference approach of the PSNR is unable to do.

Combining this with the previous two steps leads to the proposed quality metric Q_d defined as:

$$Q_d = \text{MSSIM} \left(I_d(m, n) \cdot T_d(m, n), \hat{I}_d(m, n) \cdot T_d(m, n) \right), \quad (3)$$

where $I_d(m, n)$ and $\hat{I}_d(m, n)$ are the virtual images synthesized from the original and coded II respectively.

3. EXPERIMENTAL SETUP

Four II-based 3D-images were synthesized using a II-camera model with a pinhole lenslet approximation [7]. To simulate thermal noise in the camera pixel sensor, normally distributed noise $N(\mu = 0, \sigma = 1)$ was added to each of the three 8 bit color components. In Table 1, details about the II-camera

Table 1. Experiment setup

Parameter	Value
Number of lenslets (EI)- $K \times L$	64×64
EI resolution - $U \times V$ [pixels]	64×64
Lenslet focal length - f [mm]	0.73
Lenslet pitch - Δ [mm]	0.39
Pixel sensor resolution - $M \times N$ [pixels]	4096×4096
Pixel sensor size [m^2]	25×25
Pixel pitch - δ [μm]	6.1

are summarized. The image capturing pixel sensor is in pair with e.g. the CMOS 16MP sensor used in Canon’s digital SLR camera EOS-1Ds Mark II. When evaluating the metric’s depth dependency, three coding schemes were studied: two H.264-based pseudo video coding schemes and JPEG 2000 [8]. A bit rate of 0.09 bpp where used to distinctly bring out the coding schemes’ characteristic artifacts.

4. RESULTS

Figure 4 shows the four 3D-images that have been coded with three different coding schemes. The studied depth layers are distributed approximately logarithmic throughout the scenes. The coding artifacts of coding scheme #1 shows a pronounced low-pass characteristic throughout all depths whereas coding scheme #2 tend to have less distortion at low d . Out of the three schemes, coding scheme #3 shows the most apparent and severe distortion for all images – excluding Twins.

Figure 5 collects the results from evaluating these images using the proposed metric. Note the significant dip in Apples at $d = 60$ for coding scheme #3. This depth layer corresponds to the middle apple, which can be seen in Figure 4 to also be severely affected by coding scheme #3. Another observation that can be made is the single sharp dip in quality for Car which is in line with the shallow depth range occupied by objects in that scene. Cuboid on the other hand has a wider quality reduction corresponding to the wider depth range of relatively large number of objects. In Twins, the quality dip at $d = 25$ corresponds in part to pixels forming the left woman’s face. The high frequency content of this region, together with the homogenous smearing character of coding scheme #2, results in a more severe quality reduction compared to the other schemes. A result that is in line with what can be seen when inspecting Twins in Figure 4.

5. CONCLUSION

We have presented an objective quality metric for II-based 3D-images that explicitly captures how a lossy coding scheme distributes the coding artifacts within the depth range of the 3D-scene. Next to single valued metrics, and angle dependent

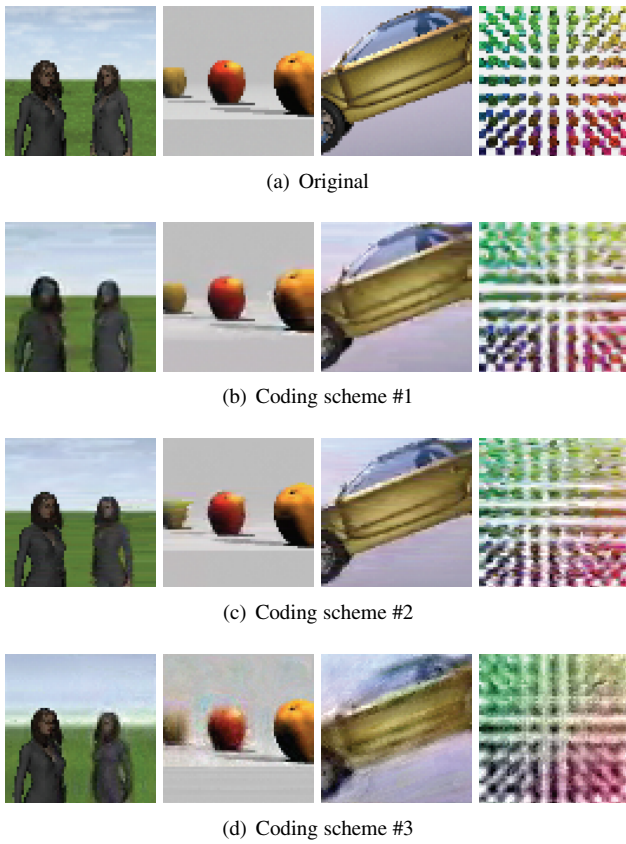


Fig. 4. Coding artifacts at $r = 0.09$ *bpp* (set low to provoke distortion) for the four reference scenes. The top images in (a) are synthesized from the original uncoded 3D-images whereas (b) – (d) correspond to three different coding schemes

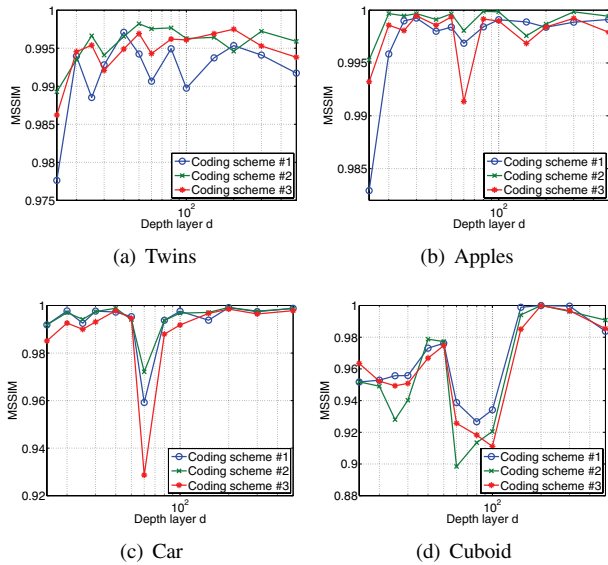


Fig. 5. Proposed quality metric evaluated on the II-based 3D-images (a) Twins, (b) Apples, (c) Car and (d) Cuboid shown in Figure 4.

metrics, this depth dependent metric is a supplementary tool when evaluating II-based 3D-image coding schemes.

This 1D quality measure shows good correlation with what can be perceived when visually inspecting the introduced coding artifacts. Still, the keen reader recognizes that the degrees of freedom in the metric's subsystems allows for further optimizations. Removing the restriction on virtual camera position can extend the metric to include viewing angle dependence. Despite being a straightforward extension it must be balanced against the complexity of a metric that then approaches the dimensionality of the signal itself.

Future work includes evaluating other depth map algorithms as this stage has a direct affect on the depth discrimination of the metric. Also, to speed up the base and reference image rendering for large number of depth layers, Fourier Slice Photography will be investigated [9]. Layer dependent normalization of the MSSIM will also be studied to increase the metric's swing and thereby its sensitivity to small scale artifacts.

6. REFERENCES

- [1] N. P. Sgouros, A. G. Andreou, M. S. Sangriotis, P. G. Papageorgas, D. M. Maroulis, and N. G. Theofanous, "Compression of ip images for autostereoscopic imaging applications," in *Proceedings of ISPA03*, September 2003, vol. 1, pp. 223 – 227.
- [2] Sekwon Yeom, Adrian Stern, and Bahram Javidi, "Compression of 3d color integral images," *Optics Express*, vol. 12, no. 8, pp. 1632 – 1642, April 2004.
- [3] Matthew C. Forman, Neil Davies, and Malcolm McCormick, "Objective quality measurement of integral 3d images," in *Proceedings of SPIE Vol. 4660 Stereoscopic Displays and Virtual Reality Systems IX*, 2002.
- [4] Roger Olsson, Mårten Sjöström, and Youzhi Xu, "Evaluation of combined pre-processing and h.264-compression schemes for 3d integral images," in *Proceedings of Electronic Imaging - VCIP*. IS&T/SPIE, January 2007.
- [5] Keita Takahashi and Takeshi Naemura, "Unstructured light field rendering using on-the-fly focus measurements," in *IEEE International Conference on Multimedia and Expo*. IEEE, July 2005.
- [6] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600 – 612, April 2004.
- [7] Roger Olsson and Youzhi Xu, "A ray-tracing based simulation environment for generating integral imaging source material," in *Proceedings of Radio Vetenskap och Kommunikation*, June 2005, pp. 663 – 666.
- [8] Roger Olsson, Mårten Sjöström, and Youzhi Xu, "A combined pre-processing and h.264-compression scheme for 3d integral images," in *Proceedings of ICIP 2006*. October 2006, pp. 513 – 516, IEEE.
- [9] Ren Ng, "Fourier slice photography," in *Proceedings of ACM SIGGRAPH 2005*, July 2005, vol. 24, pp. 735 – 744.