

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

Roger Olsson and Mårten Sjöström. A novel quality metric for evaluating depth distribution of artifacts in coded still 3D images. In *Proceedings of Stereoscopic Display and Application XCIX*, SPIE, Vol. 6803, San Jose (CA), USA, January, 2008.

<http://dx.doi.org/10.1117/12.766256>

© Copyright 2008 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.

# A novel quality metric for evaluating depth distribution of artifacts in coded 3D images

Roger Olsson and Mårten Sjöström

Dept. 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 different types of light field based 3D-images, e.g. integral imaging or multi-view. The metric results in a single accumulated quality value for the whole 3D-image. Evaluating single views – seen from specific viewing angles – gives a quality matrix that present the 3D-image quality as a function of viewing angle. However, these two approaches do not capture all aspects of the induced distortion in a coded 3D-image. We have previously shown coding schemes of similar kind for which coding artifacts are distributed differently with respect to the 3D-image's depth. In this paper we propose a novel metric that captures the depth distribution of coding-induced distortion. Each element in the resulting quality vector corresponds to the quality at a specific depth. First we introduce the proposed full-reference metric and the operations on which it is based. Second, the experimental setup is presented. Finally, the metric is evaluated on a set of differently coded 3D-images and the results are compared, both with previously proposed quality metrics and with visual inspection.

**Keywords:**

## 1. INTRODUCTION

Three-dimensional (3D) images and 3D-video based on multi-view contains large amounts of data compared to their stereoscopic counterparts. An increasing number of pixels and views are required to produce realistic 3D-properties such as increased depth range and wide range of motion parallax. For 3D-images based on integral imaging (II), the requirements are even greater due to the additional vertical parallax and finer granularity of the motion parallax. Fortunately, the additional data required for 3D images is to a large extent redundant. The captured pictures from neighboring cameras in a multi-view camera setup show high correspondence due to the cameras similar positions and orientation. Similar redundancies are present in neighboring elementary images (EI), i.e. the subsets of the captured II-picture projected beneath each lens in the lens array. This redundancy property is illustrated in Figure 1 where a synthesized II-picture is shown in subfigure (a) together with a zoomed-in portion of the II-picture containing  $6 \times 6$  EIs in subfigure (b). A stereoscopic image pair, synthesized from the coded 3D-image, is shown in Figure 2 (a). It gives an overview of the scene content, here containing three apples located at different distances from the camera. Hence, coding is a beneficial operation for many 3D-image applications, even if not required. Given the similarities between multi-view and II, the latter will be the focus of this paper due to its higher data requirement. Through-out the rest of the paper 3D-image and II-picture will be used interchangeably and refer to the same signal type.

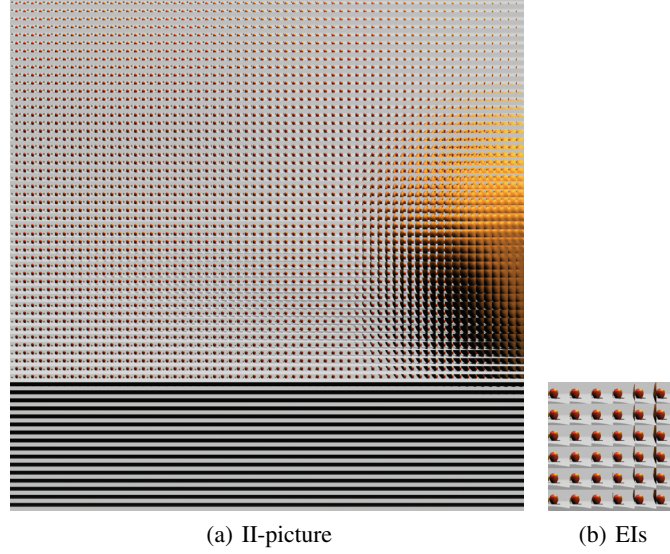
Selecting an objective quality metric is a vital step when evaluating any lossy 3D-image coding scheme that introduce distortion, especially when new coding schemes are proposed. Still, this selection is frequently left out without an explicit motivation. Instead general conventions from the 2D coding field are adopted, which involve applying the Peak Signal to Noise Ratio (PSNR) metric on the full data set of the 3D-image. PSNR is well known from 2D image processing, image- and video coding and has shown good merits as a low complex evaluation method in comparable research by which using the same quality metric is a prerequisite for comparing results. However, much of the information about the introduced distortion is lost when the quality of a coded 3D-image is described using a single scalar value. For example, in our previous work we have identified coding schemes, which mainly differ in how distortion is distributed with respect to the 3D-image's depth.<sup>1</sup> Such a depth-dependent distortion cannot be measured by a single scalar.

---

Further author information: (Send correspondence to Roger Olsson)

Roger Olsson: E-mail: Roger.Olsson@miun.se, Telephone: + (46) 60 14 86 98

Mårten Sjöström: E-mail: Marten.Sjostrom@miun.se, Telephone: + (46) 60 14 88 36



**Figure 1.** An virtual scene containing three apples which is captured in (a) a synthesized II-picture, where (b) shows a zoomed-in subset of  $6 \times 6$  EIs. An overview of the scene is shown in Figure 2 (a) where a stereoscopic image pair is synthesized from the II-picture.

In this paper we propose a novel quality metric that extends our previous work on a metric for evaluating the depth distribution of coding-induced distortion.<sup>2</sup> The metric aims to be a complement to single value metrics, and so be a useful tool when designing 3D-image coding schemes.

The structure of the paper is as follows. In Section 2 we give a review of the information deficiency imposed by present quality metrics. Section 3 gives a detailed description of the proposed quality metric and its constituting parts. The evaluation setup that is used to evaluate the metric is presented in Section 4, and Section 5 presents the results. Finally, a conclusion ends the paper.

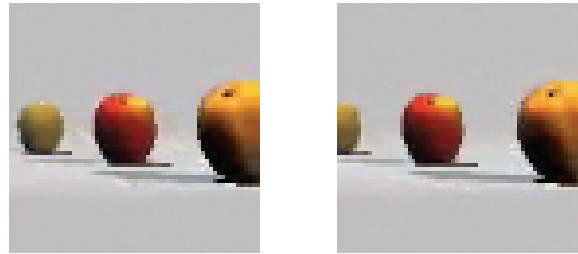
## 2. PREVIOUS WORK

The single most adopted method to assess quality in 3D-images is by globally applying PSNR to the complete dataset of pixels.<sup>3-8</sup> This has the advantage of producing a result that in a single scalar value aggregates the quality of the complete data set of the 3D-image. Hence, it lends itself well to comparative studies where a single measurement using the metric also give rise to a single value. However, this simplicity is also a drawback when it comes to gaining detailed insight into the distribution of introduced distortion. For obvious reasons it is impossible to discern local properties within a 3D-image from a single scalar value. This is illustrated in Fig. 2 where the two 3D-images (from which the stereoscopic image pairs are synthesized) have the same global PSNR (28dB) yet still show very different coding artifacts due to different coding schemes.

Forman et. al<sup>9</sup> proposed a viewing-angle-dependent quality metric to give a more informative result. Their proposed metric produced a vector for every measurement instance due to the horizontal parallax only (HPO) 3D-images used. Each element in the vector represents the quality of the 3D-image perceived by a virtual viewer from a specific viewing angle at an infinite viewing distance from the 3D-display. Figure 3 illustrates this quality metric applied to the coded 3D-image being the origin of the synthesized stereo pairs in Figure 2 (a) and (b). Only the horizontal component of the full parallax 3D-image of this work was evaluated, since the metric was defined for HPO 3D-images. Note that the variations in PSNR over the viewing range is significant compared to the average PSNR. In addition, a comparison of Figure 3 (a) and (b) reveals the lack of tendency in PSNR over the viewing angles. Hence, a greater understanding of the introduced coding artifacts can be achieved by using the viewing angle dependent quality metric. Still, the apparent difference in depth distribution of coding artifacts in Figure 2 (a) and (b) cannot be discerned from Figure 3.

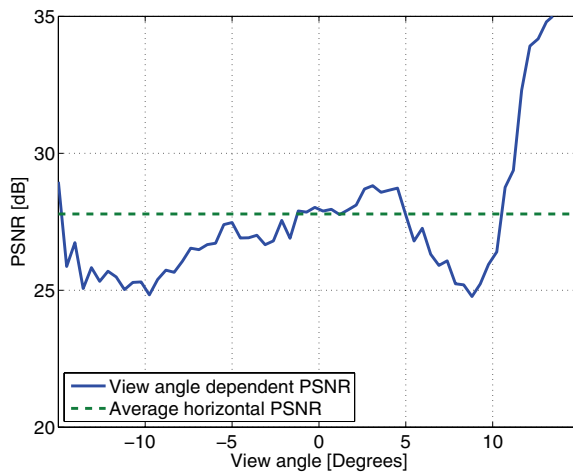


(a) Coding scheme #1 at 0.08 bpp

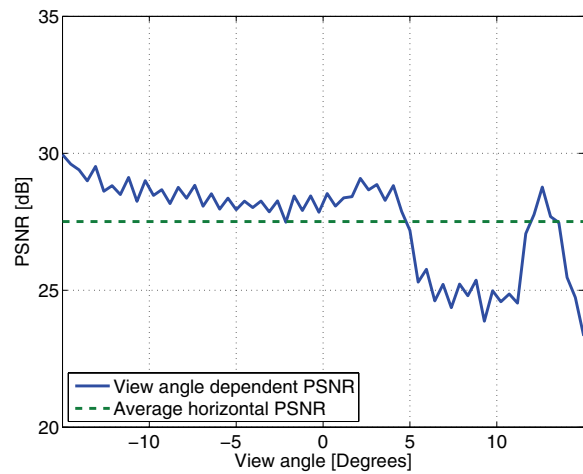


(b) Coding scheme #2 at 0.14 bpp

**Figure 2.** Coded images that share the same global PSNR = 28dB, yet show different distortion properties depending on used coding scheme. (a) pre-processing and H.264-based coding scheme #1 (0.08 bpp) and (b) JPEG 2000 (0.35 bpp). The synthesized right and left views are positioned for cross-eyed viewing.

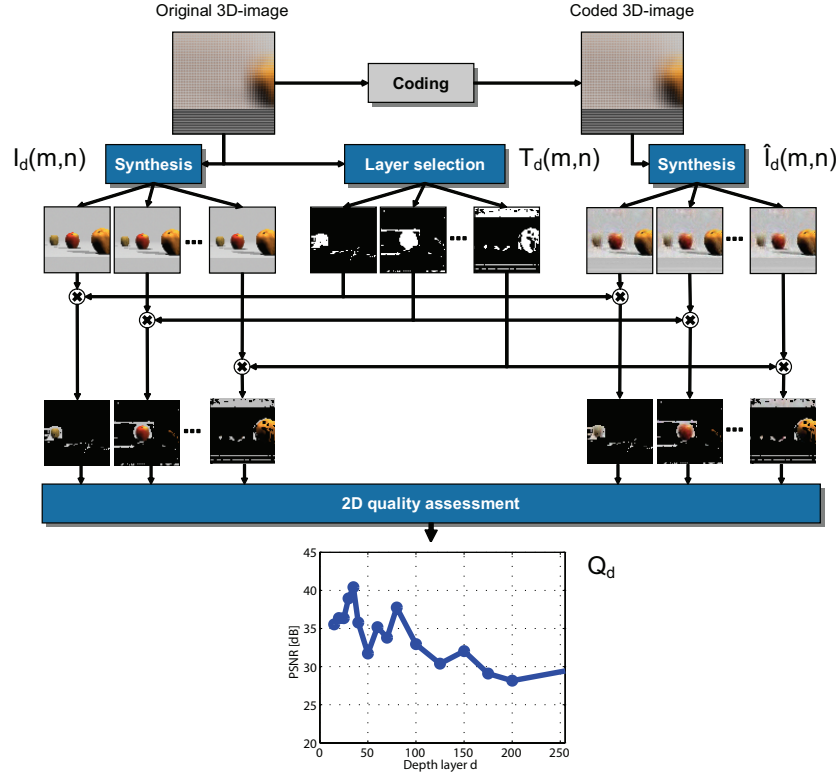


(a) Coding scheme #1



(b) Coding scheme #2

**Figure 3.** Viewing angle dependent quality metric applied to a 3D-image coding using (a) pre-processing and H.264-based coding scheme #1 and (b) pre-processing and H.264-based coding scheme #2.



**Figure 4.** Block diagram for the proposed quality metric.

### 3. PROPOSED QUALITY METRIC

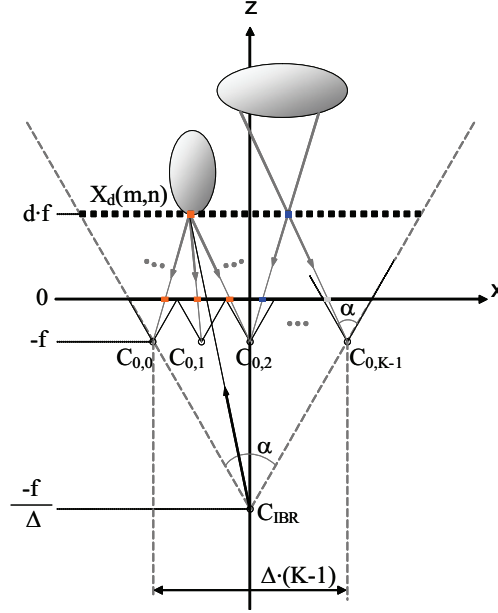
This section first gives a general description of the proposed quality metric, which is then followed by a detailed description of the three constituting operations of the metric.

The proposed metric is also a full-reference metric, i.e. it requires access to both the coded and the original 3D-image for its operation. From these two 3D-images:

1. a set of 2D-image pairs are synthesized using image based rendering (IBR) with focal plane located at different distances from the virtual camera,
2. pixels that can be discerned to belong to a specific distance (depth layer) are identified, and
3. image pairs are constructed (for each depth layer) and are then evaluated using a 2D quality metric.

These steps results in a quality value per depth layer, which define a 3D-quality metric able to measure distortion at different depths within the 3D-image. Figure 4 illustrates the operations that constitutes the metric. These process enumerated above and presented in the block diagram can be modularized into the operations:

- Synthesis using IBR
- Layer selection using depth estimation
- Quality assessment using a 2D quality metric



**Figure 5.** Geometry of the virtual camera placement relative to the II-camera's lenslets. Only the  $xz$ -plane with the first row of lenslets ( $C_{0,k}$ ) is shown to simplify the figure.

These are performed independently of each other, where the output of the first forms the input to the second and so on. Hence, the modularity of the metric allows it to be parameterized in many ways. The following description of the metric is based on a single parametrization but alternative methods for performing each operation will be discussed when appropriate.

### 3.1. Synthesis of virtual views

Synthesis of virtual views have two objectives in the proposed quality metric:

1. Aiding the depth layer selection
2. Producing input for the 2D quality metric

In both cases the novel 2D virtual views are synthesized from the 3D-image data using a plane-model IBR technique, i.e no assumptions are made regarding scene content in order to enhance the rendering quality of each view.

The positioning of the virtual camera  $C_{IBR}$ , from which the views are rendered, can be placed arbitrarily. However, by constraining the  $xyz$ -position to  $[0, 0, \frac{-f}{\Delta}]$ , where  $f$  and  $\Delta$  is the focal length and pitch of each lens respectively, two favorable properties are achieved. First, any angle dependency is eliminated from the metric, which is primarily a depth-dependent metric (a more complex metric with higher dimensionality can be achieved by positioning  $C_{IBR}$  at several other locations in addition to the optical  $z$ -axis). Second, the specific distance to the image plane of the II-camera ( $\frac{-f}{\Delta}$ ) ensures that the field-of-view of the virtual camera and II-camera are equal. This results in an efficient use of the 3D-image data as all EIs contribute to the synthesized views. Figure 5 illustrates the geometry of the IBR technique, especially how the focal plane of  $C_{IBR}$  is positioned at a specific depth layer  $d$ . As the figure illustrates, depth layer  $d$  (which is parallel to the  $xy$ -plane) correspond to a physical distance from the II-camera that is a multiple of the lenslet focal length  $f$ . Hence, an object at distance layer  $d = 100$  is positioned  $100 \cdot f$  meters from the image plane of the II-camera.

When synthesizing the image pair, it is vital not to destroy any coding artifacts present as they will be used as input to the 2D quality metric. Often interpolation is used in IBR to enhance the quality of the final images. However, when producing these image pairs all interpolation must be avoided as it would have a low-pass filtering affect on the distortion, i.e. it would smear away any coding artifacts and consequently influence the measurements. Therefore, only one EI pixel

contributes to the color when calculating the pixel color values of the image pairs  $I_d$  (view from original 3D-image) and  $\hat{I}_d$  (view from coded 3D-image). Both views share the same focal plane, which is located at depth layer  $d$ . In Figure 5, this implies that the color of a image pixel  $X_d(m, n)$  (marked red) is taken to be the color of the pixel in the EI covered by  $C_{0,1}$  alone. If instead the pixel color value would be calculated from the sum of neighboring EI's pixels, the operation would average any distortion that is not shared by all EIs.

When synthesizing views for the depth layer estimation is instead beneficial to average over a larger number of EIs when calculating the pixel color values. This effectively increases the aperture of  $C_{IBR}$  and a larger aperture results in a shorter depth of field, which is beneficial for the discriminating between depth layers.

Regardless of how many EIs contribute to the final pixel color value, a selection must be made as to which pixel to use from each EI. This selection is made by projecting the position of the resulting virtual image pixel onto the xy-plane using the projection matrix of the EI, as shown to the left of Figure 5. Projecting onto the xy-plane might result in fractional pixel positions within the EI. Nearest neighbor interpolation is used to address this for the image pair synthesis whereas bilinear interpolation is adopted for the depth layer estimation.

### 3.2. Depth layer estimation and selection

Only a subset of the pixels will correspond to objects intersected by a specific depth layer when synthesizing a view with a focal plane at that depth layer. Hence, a selection must be performed where pixels corresponding to objects outside the depth layer are 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.<sup>10</sup> In their work 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 5) will have similar pixel color values from contributing EIs. Other image pixels (e.g. the blue pixel in Figure 5) will be an average of objects *outside* the depth layer and the colors will vary. Evaluating the focus measure at different depths allows for a pixel-resolution depth map to be constructed by:

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

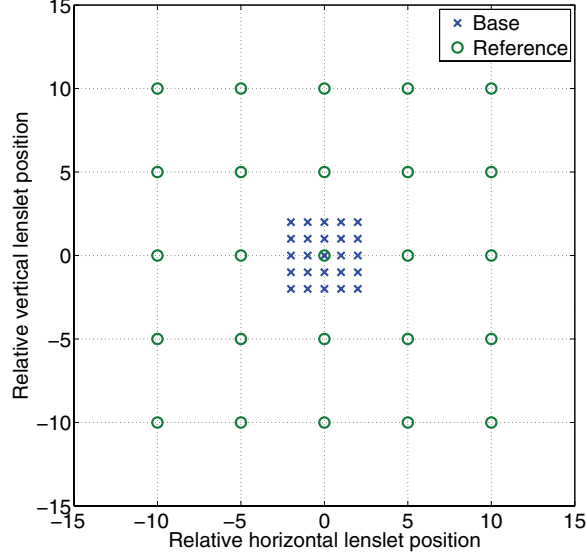
where  $h$  is a filter kernel that is used in a convolution step (operator  $*$ ) to enhance the result, as described in.<sup>10</sup>

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 set increases the reliability of the focus measure as the depth of field is reduced, which consequently enhances the accuracy of the depth map. Compared to the original focus measure definition, we therefore increase the number of EIs contributing to each image pixel from 2 to 25. Increasing the set of EIs have the cost of increasing the synthesis time and hence also the time to calculate the proposed quality metric. Figure 6 shows the EIs patterns used to select which neighboring EIs that contributes to the pixel color values 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:

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

### 3.3. 2D quality metric

Different aspects should be considered when selecting which 2D quality metric to apply on the extracted depth layers: low complexity, efficient computation, easily interpreted result etc. A good metric should capture the complexity of the human visual system (HVS) and produce results that correspond to those achieved in subjective tests conducted with viewer groups. An example of such a metric is the Mean Structural Similarity index (MSSIM), which has shown high prediction consistency with subjective test results.<sup>11</sup> However, adopting a quality metric consistent with subjective test results is only valid if the image to be evaluated is a directly viewable form. The masked depth layer images however are in a form that will not be viewed per se; the viewer will see the combined set when viewing the 3D-image on a 3D-display. Hence, the consistency with subjective tests results can be relaxed here. As a result, we suggest instead a metric that adopts PSNR as the 2D quality metric. Its well-known value-range is a strong advantage. It can, however, not be directly compared with applying the PSNR globally to the whole 3D-image when used in the context of depth layer images.



**Figure 6.** Relative lenslet positions that are used to derive a specific image pixel for the base- and reference virtual image respectively.

The definition of the full color PSNR in dB for an uncoded image  $X$  and its coded counterpart  $\hat{X}$  is:

$$PSNR(X, \hat{X}) = 20 \cdot \log_{10} \left( \frac{255}{\sqrt{MSE}} \right) \quad (3)$$

where

$$MSE = \frac{1}{3 \cdot M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left| X(m, n) - \hat{X}(m, n) \right|^2 \quad (4)$$

The average operation performed in Equation (4) presumes that all  $M \cdot N$  pixels in the two images contain valid color values. However, this is not true when  $X = I_d \cdot T_d$  and  $\hat{X} = \hat{I}_d \cdot T_d$ . Only  $M_d \cdot N_d = \sum_{\forall m, n} T_d(m, n)$  of the  $M \cdot N$  pixels

belong to depth layer  $d$ . Normalizing using  $M \cdot N$  gives equal weight to each layer, regardless of its size in pixels. That is, a layer with a small number of severely distorted pixels may give PSNR value equal to a layer with a large number of pixels showing minor coding artifacts. Normalizing using  $M_d \cdot N_d$  instead gives equal importance to each depth layer, regardless of size. The latter normalization results in increased discernibility of coding artifacts and is therefore used henceforth.

### 3.4. Operations combined

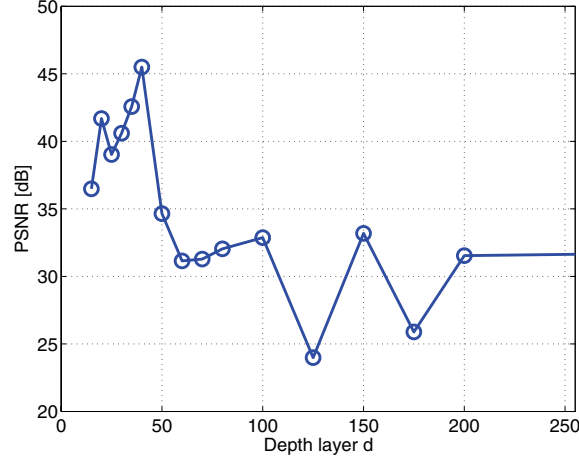
Combining the above steps gives the proposed quality metric  $Q_d$  defined as:

$$Q_d = PSNR \left( I_d \cdot T_d, \hat{I}_d \cdot T_d \right) \quad (5)$$

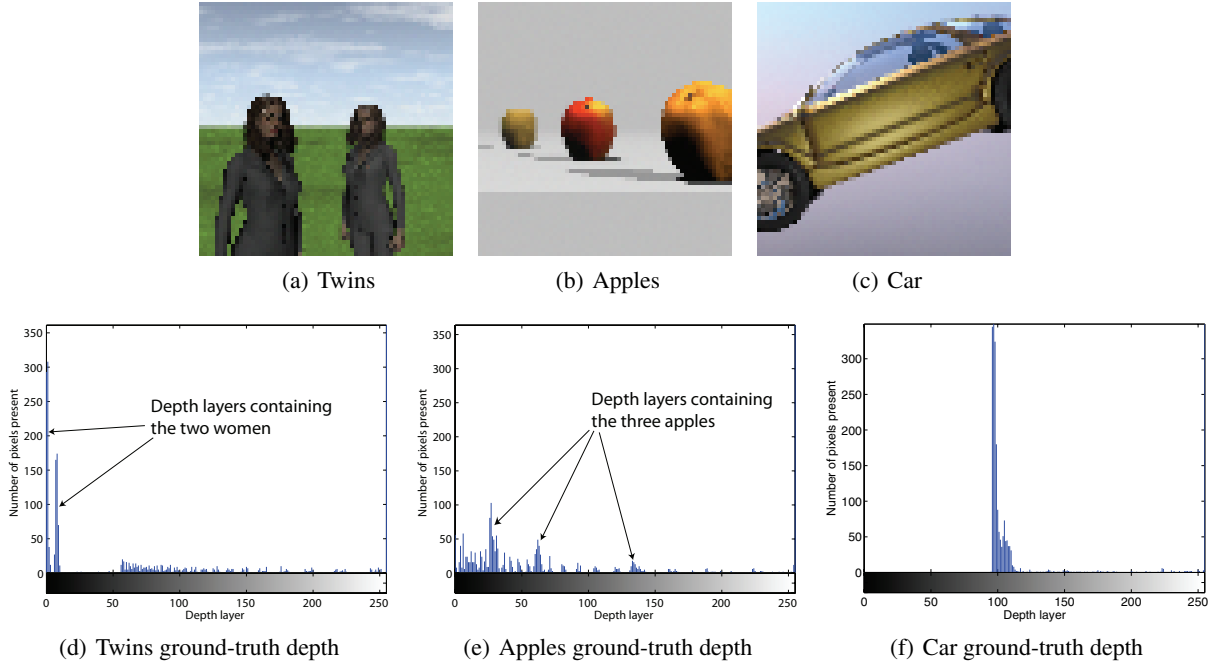
A linear (as well as a logarithmic) distribution of  $d$  can be adopted if no a priori knowledge is available about the 3D-image's depth content. An approximate depth distribution gained from a pre-processing depth estimation step could also be utilized. The selection of which depth layer to evaluate is, however, outside the scope of the metric.

Figure 7 shows the quality metric applied to the coded 3D-image being the source for the stereoscopic pair in Figure 2 (a). From comparing the graph with the image pair in Figure 2 (a), the proposed metric is able to discern the difference in depth distribution of the induced coding artifacts. Coding scheme #1 produces higher values in  $Q_d$  for nearby objects (low  $d$ ), at the expense of high distortion on far objects. This property is reflected in both the stereo pair as well as in the proposed metric. The metric also identified the severely distorted apple farthest from the camera with its minimum PSNR below 25dB at  $d = 125$ . Further analysis of the proposed metric is left for Section 5.





**Figure 7.** Proposed quality metric applied to a 3D-image coded using the H.264-based coding schemes #1.



**Figure 8.** The three 3D-images Twins, Apples and Car and their respective perspective projections and ground-truth depth distributions.

#### 4. EVALUATION SETUP

The metric was evaluated using four coded versions of the 3D-images Twins, Apples and Car. The images were rendered using a II-camera model with pinhole lenslet approximation and normally distributed noise  $N(\mu = 0, \sigma = 1)$  added to each of the three 8 bit color components.<sup>12</sup> Perspective projections of the images are shown in Figure 8 combined with ground-truth depth distributions extracted from rendering. The II-camera model's pixel sensor was set to be in pair with e.g. the CMOS 16MP sensor used in Canon's digital SLR camera EOS-1Ds Mark II. See Table 1 for details regarding the II-camera model and the 3D-image properties. The coding schemes used was two variants of combined pre-processing and H.264-coding schemes,<sup>4,7</sup> JPEG 2000 applied to the full II-picture and finally the JPEG 2000 3D (Part 10 - JP3D) applied to the stack of EIs.<sup>13</sup> A subset of all integer depth layers were evaluated to reduce computation time. The evaluated depth layers (indicated by circle plot symbols in the result graphs) are a compromise between depth layers with known objects and control layers with no known objects.

**Table 1.** Experiment setup

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

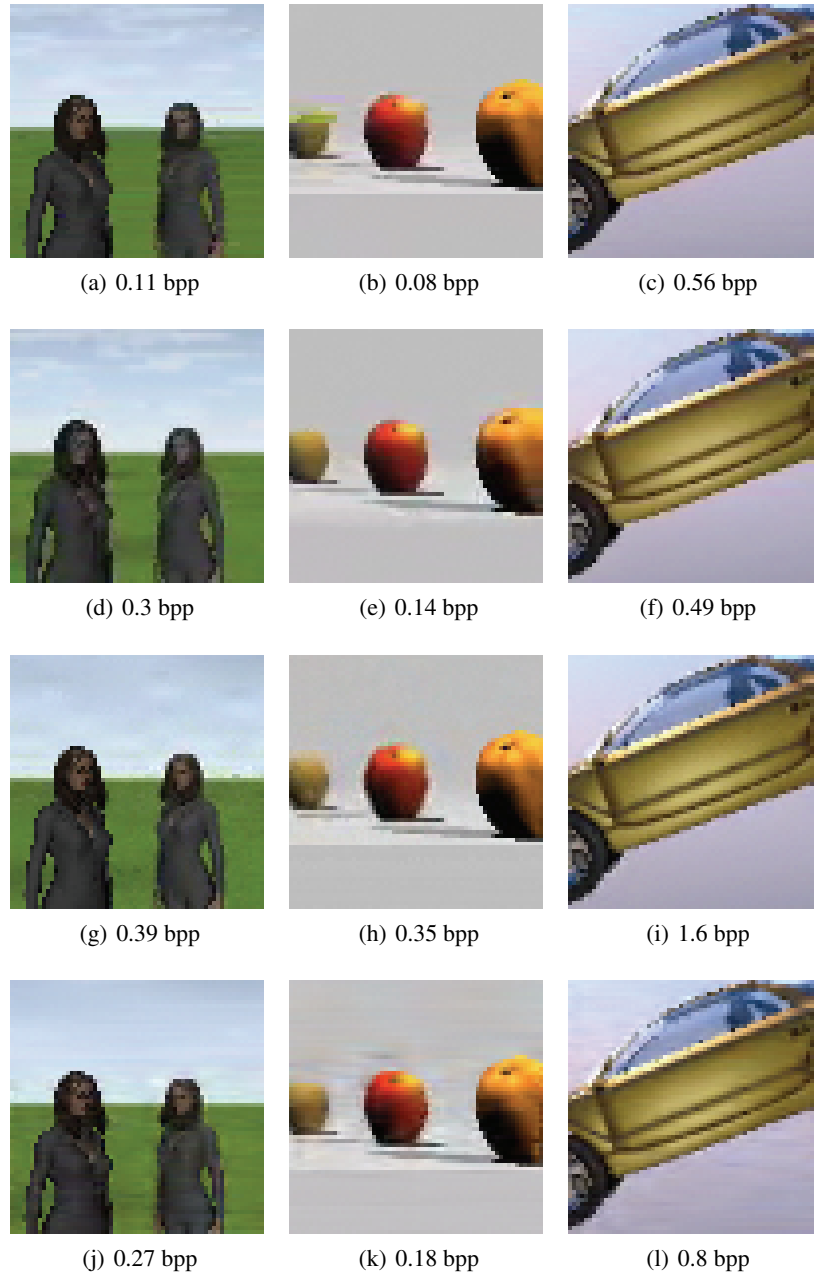
## 5. RESULTS

Figure 9 presents center views of the 3D-images coded using all coding schemes, i.e. 12 perspective projection images. These will be used for visually validating what can be inferred from the proposed quality metric results. The proposed metric for all coding methods are combine in the three subfigures of Figure 10. Figure 10 (a) has a lower depth range than subfigure (b) and (c) to facilitate examination of the effects at low depth layers  $d < 50$ , which according to Figure 8 (d) is where the interesting depth layers are located. Maintaining the number of depth layers over all 3D-image evaluations meant transferring a few from high to low  $d$ .

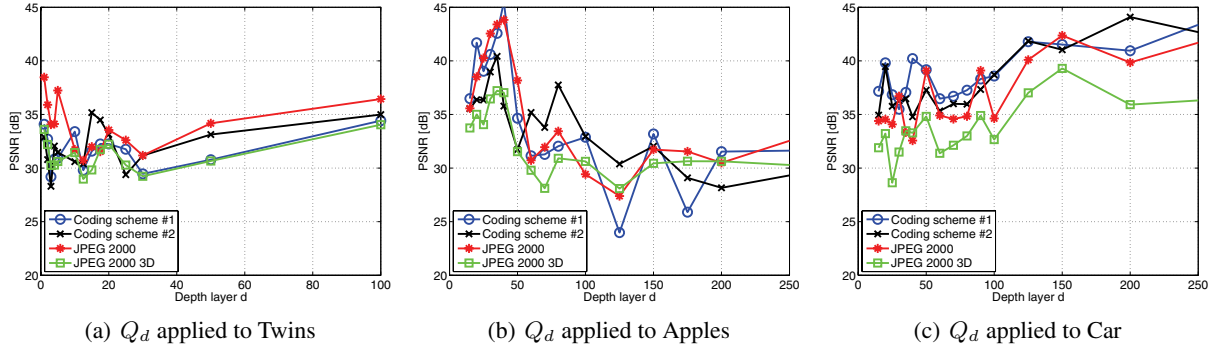
A notable property of the  $Q_d$  graphs in Figure 10 (a) is that JPEG 2000 have less distortion on the nearest woman whereas coding scheme #1 reproduce the farthest woman with less amount of coding artifacts. This is to some extent confirmed in Figure 9 (a), (d), (g) and (j). The results of 3D-image Apples (Figure 10 (b) shows peaks in  $Q_d$  at low depth layers  $d$  for coding scheme #1 and JPEG 2000, whereas  $Q_d$  is reduced for increasing depth layers  $d$ . Coding artifacts mainly located at distant objects is also verified visually in Figure 9 (a) and (g). Coding scheme #2 shows a less steep decline in  $Q_d$  for increasing depth, and the fluctuation over all evaluated depths is also lower. Figure 9 (d) shows artifacts that are in line with this analysis: the distortion is more evenly distributed throughout the 3D-image's depth than for coding scheme #1 and JPEG 2000. The 3D-image coded using JPEG 2000 3D produce an even more flattened graph. A slight increase in  $Q_d$  is present at low  $d$ , but for the large part of the depth range the metric stays approximately constant at a low lever. This result imply a number of coding artifacts spread over the 3D-image depth. The clearly visible patchy-like artifacts of Figure 9 (j) is randomly distributed at different depths due to low correlation between views, which causes insufficient binocular parallax to determine depth. The  $Q_d$  graphs of Figure 10 (c) shown no distinguishing properties. No distinguishing difference is visible when examining the center images of Figure 9 (c), (f), (i) and (l) either. The false positive identification of depth layers without known objects ( $|d - 105| > 10$ ) is mainly the result of depth estimation failure when detecting parallax in the smooth uniform-colored surfaces of Car.

Analyzing rate-distortion properties is a vital aspect of evaluating coding schemes. A good coding scheme should degrade the quality gracefully when the rate is reduced. Each measurement using the proposed metric gives rise to a graph per bit rate tested, as shown in Figure 10. One way to use the metric, without reducing the information contained, is to show the result  $Q_d$  as an image – each pixel containing a PSNR value at a given rate and depth. Such examples are shown in Figure 11, which contains coded versions of 3D-image Apples. This specific representation form positions the rate-distortion image pixels equidistantly – both horizontally and vertically – even though bit rate nor depth layers are uniformly distributed. An exact positional correspondence could be achieved by resampling. The presentation format of Figure 11 gives a general view while still allowing for detailed analysis. An example is the result of coding scheme #2 in Figure 11 (a), which succeeds in producing high  $Q_d$  throughout the whole depth range for high bit rates. It also shows a smooth degradation when the rate is reduced. The JPEG 2000 coding of the 3D-image (Figure 11 (b)) on the other hand does not share these characteristics. Instead, objects further away are subject to coding artifacts even at high bit rates, and even more so when the rate is reduced.

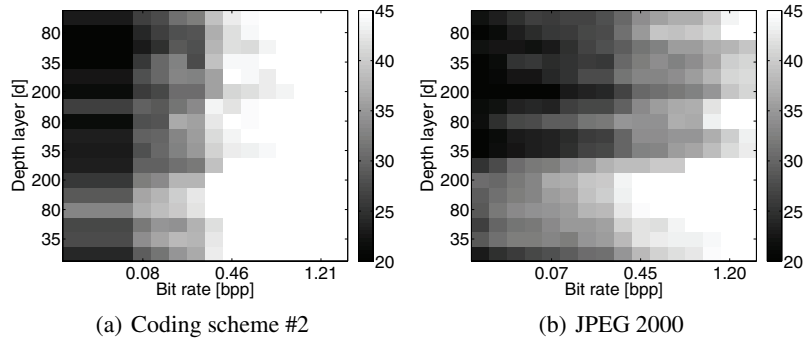
Classical rate-distortion graphs can also be produced from the metric data. Mean, min, max, standard deviation etc. are operators that reduce the dimensionality of the metric to a scalar value, which might be feasible for single valued analysis that considers the depth dependency of the distortion. A set of such graphs are shown in Figure 12 where for each bit rate the mean, minimum and maximum  $Q_d$  is jointly displayed. An observation to be made from the rate-distortion graphs is the difference in spread of distortion levels that the coding schemes give rise to. For example, despite the similar mean



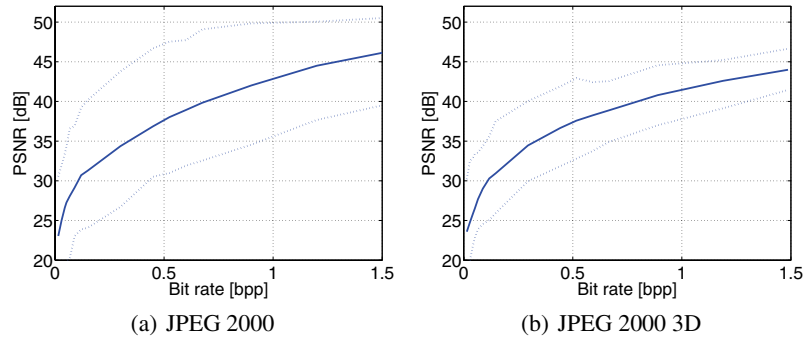
**Figure 9.** Center views from the 3D-images Twins, Apples and Car coded at a global PSNR = 28 dB using (a)-(c) Coding scheme #1, (d)-(f) Coding scheme #2, (g)-(i) JPEG 2000 and (j)-(l) JPEG 2000 3D. The bit rate used to produce the global PSNR is presented beneath each center view.



**Figure 10.** Proposed depth-dependent metric  $Q_d$  applied to the 3D-images (a) Twins, (b) Apples and (c) Car. Each image coded using four coding methods.



**Figure 11.** Rate-distortion images produced from 3D-image Apples coded with (a) coding scheme #2 and (b) JPEG 2000.



**Figure 12.** 3D-image Apples analyzed using mean  $Q_d$  (solid line) with accompanying minimum (lower dashed line) and maximum (upper dashed line) values for (a) JPEG 2000 and (b) JPEG 2000 3D.

$Q_d$  between JPEG 2000 and JPEG 2000 3D they differ significantly with respect to the  $\min(Q_d)$  and  $\max(Q_d)$  at each bit rate. JPEG 2000 3D shows a close collected  $Q_d$  value range whereas JPEG 2000 exhibits a more spread out  $Q_d$ . For example,  $\max(Q_d) - \min(Q_d) \approx 15dB$  at  $0.5bpp$ . Worth remembering though is that the majority of the information extracted using the proposed quality is lost when its results are reduced like this to a small number of scalar values.

## 6. CONCLUSIONS

We have presented an objective quality metric for 3D-images, which can discern the depth distribution of coding induced distortion. The metric can act as a supplementary tool when evaluating 3D-image coding schemes in addition to single valued global metrics and angle dependent quality metrics. New aspects of coding artifacts in 3D-images may be revealed using the metric, both in the metrics original form as well as after post-processing of the result. For example, rate-distortion images can show coding artifact as a function of both rate and depth giving a broader understanding of the effect of coding.

Future work includes investigating other parameterizations of the metric where for example other depth estimation algorithms is of special interest.

## ACKNOWLEDGMENTS

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

## REFERENCES

1. R. Olsson, M. Sjöström, and Y. Xu, "Evaluation of combined pre-processing and h.264-compression schemes for 3d integral images," in *Proceedings of Electronic Imaging - VCIP*, IS&T/SPIE, (San Jose (CA), USA), January 2007.
2. R. Olsson and M. Sjöström, "A depth dependent quality metric for evaluation of coded integral imaging based 3d-images," in *3DTV Conference*, IEEE, EURASIP and MPEG-IF, (Kos, Greece), May 2007.
3. 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*, **1**, pp. 223 – 227, September 2003.
4. S. Yeom, A. Stern, and B. Javidi, "Compression of 3d color integral images," *Optics Express* **12**, pp. 1632 – 1642, April 2004.
5. M. Tanimoto, "Free viewpoint television - ftv," in *Picture Coding Symposium 2004*, 2004.
6. F. Shao, G. Jiang, K. Chen, M. Yu, and T.-Y. Choi, "Ray-space data compression based on prediction technique," in *International Conference on Computer Graphics, Imaging and Vision: New Trends*, pp. 347 – 350, July 2005.
7. R. Olsson, M. Sjöström, and Y. Xu, "A combined pre-processing and h.264-compression scheme for 3d integral images," in *Proceedings of ICIP 2006*, pp. 513 – 516, IEEE, (Atlanta (GA), USA), October 2006.
8. M. Zwicker, A. Vetro, S. Yea, W. Matusik, H. Pfister, and F. Durand, "Resampling, antialiasing, and compression in multiview 3-d displays," *Signal Processing Magazine, IEEE* **24**, pp. 88–96, November 2007.
9. M. C. Forman, N. Davies, and M. McCormick, "Objective quality measurement of integral 3d images," in *Proceedings of SPIE Vol. 4660 Stereoscopic Displays and Virtual Reality Systems IX*, 2002.
10. K. Takahashi and T. Naemura, "Unstructured light field rendering using on-the-fly focus measurements," in *IEEE International Conference on Multimedia and Expo*, IEEE, July 2005.
11. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**, pp. 600 – 612, April 2004.
12. R. Olsson and Y. Xu, "A ray-tracing based simulation environment for generating integral imaging source material," in *Proceedings of RadioVetenskap och Kommunikation*, pp. 663 – 666, (Linköping, Sweden), June 2005.
13. T. Bruylants, A. Munteanu, A. Alecu, R. Deklerck, and P. Schelkens, "Volumetric image compression with jpeg2000," in *Biomedical Optics & Medical Imaging, SPIE Newsroom*, SPIE, 2007.