Evaluation of a Combined Pre-processing and H.264-compression Scheme for 3D Integral Images

Roger Olsson, Mårten Sjöström and Youzhi Xu

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

ABSTRACT

To provide sufficient 3D-depth fidelity, integral imaging (II) requires an increase in spatial resolution of several orders of magnitude from today's 2D images. We have recently proposed a pre-processing and compression scheme for still II-frames based on forming a pseudo video sequence (PVS) from sub images (SI), which is later coded using the H.264/MPEG-4 AVC video coding standard. The scheme has shown good performance on a set of reference images. In this paper we first investigate and present how five different ways to select the SIs when forming the PVS affect the schemes compression efficiency. We also study how the II-frame structure relates to the performance of a PVS coding scheme. Finally we examine the nature of the coding artifacts which are specific to the evaluated PVS-schemes. We can conclude that for all except the most complex reference image, all evaluated SI selection orders significantly outperforms JPEG 2000 where compression ratios of up to 342:1, while still keeping PSNR > 30 dB, is achieved. We can also confirm that when selecting PVS-scheme, the scheme which results in a higher PVS-picture resolution should be preferred to maximize compression efficiency. Our study of the coded II-frames also indicates that the SI-based PVS, contrary to other PVS schemes, tends to distribute its coding artifacts more homogenously over all 3D-scene depths.

Keywords: integral imaging, 3D-image coding, sub images, H.264

1. INTRODUCTION

For decades three-dimensional (3D) video has been pursued as the future of video communication. Integral imaging (II) is an auto-stereoscopic 3D technique which places no requirements of special glasses on the viewer.[1] However, a large increase in spatial resolution is required. Fortunately, with this increase a large portion of redundant information is introduced, allowing for the use of efficient coding schemes.

In the II-camera incoming light is relayed through an array of lenses before being captured by a pixel array. The large number of identical - but slightly translated - lenses act together as an optical multiplexer implicitly storing the scene depth into the pixel array. The set of pixels covered by each lens is called an elementary image (EI) and stores a unique low resolution projection of the scene. Together, all EIs form the complete II-frame. The II-display inverses the capturing process by the use of a similar lens array. The different views captured in the II-frame are distributed into different directions of the viewing space as shown in Fig. 1. Only two separate views are required to achieve binocular parallax and perceived scene depth from a fixed viewing point. However, to provide motion parallax, i.e. allow the viewer to move and still perceive a correct perspective of the scene, a larger number of views must be provided. Thus, the 3D depth fidelity comes with a cost: the number of pixels in the pixel array must increase proportionally to the number of distributed views. It has been shown that in order for the II's inherently discrete views to be perceived as continuous parallax, an increase - compared to the 2D pixel resolution - of more than ten times in both the vertical and horizontal direction is required.[2]

The increased data requirement of II has previously been addressed using different approaches. Both a 3D-DCT, as well as a hybrid DPCM/DCT scheme, have been used to code II-frames containing only horizontal parallax.[3] The set of EIs has been combined into a sequence that was then coded using a combination of JPEG-like intra-coding and MPEG-like inter-coding techniques.[4] A similar approach was later proposed, where II-frames with a low number of high resolution EI was coded using the MPEG-2 video coding standard.[5] However, in our previous work we have shown that forming a sequence from EIs is not feasible for all II-structures.[6] We therefore proposed a coding scheme adapted to full parallax II-frames with
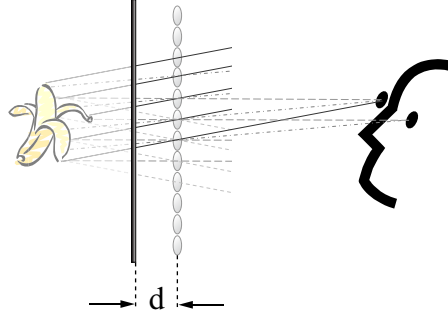

Figure 1. II-display with pixel and lens array. Rays passing a given lens correspond to an elementary image (EI). Parallel rays from different lenses correspond to a sub-image (SI). Rays converging in one eye correspond to one view image (VI).

a large number of low resolution EIs – a II-frame structure which the previous schemes fail to code efficiently.[6] Contrary to the schemes using EIs our proposed scheme transforms the II-frame into a set of sub-images (SI). This formed pseudo video sequence (PVS) is then coded using the video coding standard H.264/MPEG-4 AVC. Combined, this results in a significantly improved compression efficiency, both when compared to the EI-based PVS-scheme[4,5] as well as JPEG 2000 coding of the full II-frame.

In this paper we evaluate how parameterization of our proposed scheme affects compression efficiency as well as how it relates to the EI-based PVS-scheme.[5] We also investigate the character of the coding artifacts for the two different PVS approaches. The structure of the paper is as follows: First we define the used II-frame structure in Section 2. The proposed scheme, and how different variants of it redistribute the II-frame pixels, is elaborated on in Section 3. In Section 4, we define the three evaluation setups that are used to study the effects of the parameterization, the change in II-frame structure and the examination of PVS coding artifacts. The results from these studies are shown in Section 5 and concluding remarks are given in Section 6.

2. II-FRAME STRUCTURE

Due to the large amount of identical translated lenses that constitutes the lens array, the II-frame contains a large amount of similar projections of the scene. The aim of a pre-processing stage is to reorganize the II-frame into a form that simplifies the access to this introduced redundancy. By forming a PVS, i.e. transforming the II-frame into a form that in all essentials resembles a video sequence, enables the utilization of efficient video coding methods for compressing still II-frames. The inherent directional redundancy of the II-frame is addressed by the motion compensation of the video encoder.

We start by defining the II-frame, with a resolution of $M \times N$ pixels as

$$II\left(m, n\right), \tag{1}$$

where $m = 0, 1, ..., M - 1$ and $n = 0, 1, ..., N - 1$ are the horizontal and vertical positions of a II-frame pixel respectively. $II(m, n) = [ii_R(m, n), ii_G(m, n), ii_B(m, n)]^T$ then corresponds to the RGB-color of the m-th column and n-th row pixel.

In our study we use rectangular EIs positioned in a rectangular pattern, as illustrated in Fig. 2 (a). That is, the EI at row k and column l is defined as

$$EI_{k,l}(u, v) = II(k \cdot U + u, l \cdot V + v), \tag{2}$$

where $u = 0, 1, ..., U - 1$ and $v = 0, 1, ..., V - 1$ are the horizontal and vertical pixel positions within the EI. Thus, each of the $K \cdot L$ EIs has a resolution of $U \times V$ pixels. Other EI shapes and position patterns could be used provided that a transformation stage is introduced that bridges the gap between any non-rectangular II-frame structure and the rectangular form of the PVS-pictures.
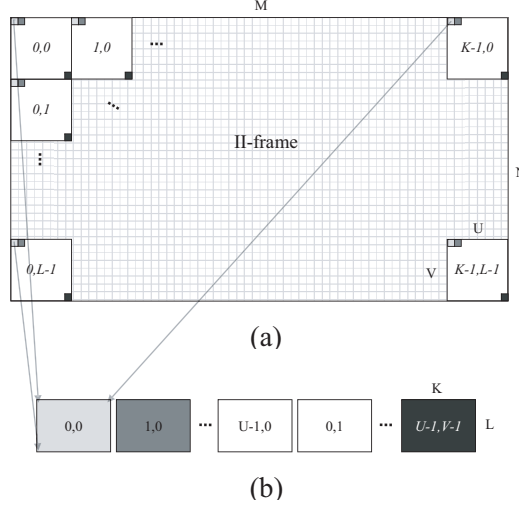
Figure 2. II-frame structure with (a) II-frame and marked EIs and (b) the forming of SIs.
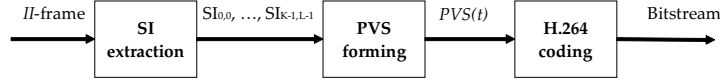
Figure 3. Block diagram of the SI-based PVS coding scheme.

3. PROPOSED SCHEME

In our proposed scheme the PVS is formed out of SIs according to the block diagram in Fig. 3. The SI is an extension of a concept used in the field of depth estimation.[7,8] Each SI comprises the subset of pixels in the II-frame that share the same relative horizontal and vertical offset to each EI center, see Fig. 2 (b). In Fig. 1, pixels that are intersected by parallel lines of the same line-style belong to the same SI. An SI is thus defined as

$$SI_{u,v}(k,l) = II(k \cdot U + u, l \cdot V + v), \tag{3}$$

where $k = 0, 1, ..., K - 1$ and $l = 0, 1, ..., L - 1$ are the horizontal and vertical pixel positions within an SI. Each of the $U \cdot V$ SIs then has a resolution of $K \times L$ pixels.

The fact that the SI is formed from parallel light rays results in it's characteristic *orthoscopic* projection property, i.e. contrary to the *perspective* projection of the EI a change in object depth does not result in a size change of the objects projection onto the SI. In Fig. 4, the II-frame, an EI and an SI is shown in (a), (b) and (c) respectively – all independently scaled for presentation purposes. Note the orthographic projection property of the SI: the projections of the two women are of the same size even though they are positioned at different depths as can be seen in the perspective projection in Fig. 6 (b).

3.1. SI-selection orders

An important parameter of our proposed PVS-scheme is the different order of selection that can be used when forming the PVS from the $U \cdot V$ SIs. The selection orders can be explicitly incorporated into a PVS-definition as

$$PVS(t) = SI_{u(t),v(t)}(k,l), \tag{4}$$

where t is the pseudo time index, or picture number, of the PVS. For the SI-based PVS $t = 0, 1, ..., U \cdot V$. The functions $u(t)$ and $v(t)$ maps pseudo time index to any given SI, i.e. choosing different functions $u(t)$ and $v(t)$ results in different
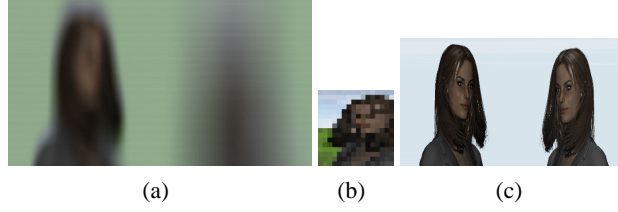
Figure 4. Example of (a) an II-frame, (b) an EI and (c) an SI. Note the low resolution of the EI and the orthographic projection property of the SI.
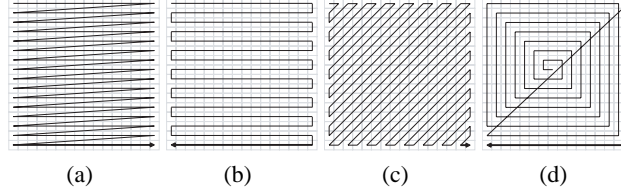
Figure 5. Selection pattern (a) row, (b) parallel, (c) zig-zag and (d) spiral.

selection orders. In this paper five selection orders for the SI-based PVS are used: row, column, parallel, zig-zag and spiral. In Fig. 5, four of these are illustrated for $U \times V = 16 \times 16$; column is excluded since it merely is the transposed row selection order.

4. EVALUATION SETUPS

In order to perform the different tests, four reference scenes – with different degrees of detail, depth and fill factor – have been defined and computer synthesized using our II-rendering tool.[9] A pinhole lens approximation was used to model the II-camera. Thermal noise in the capturing pixel array was modeled by adding independent normally distributed noise $N(\mu = 0, \sigma = 1)$ to each 24 bit pixel's red, green and blue component. The reference scenes, and their respective mix of characteristics, are summarized in Table 1 The four scenes are shown in Fig. 6.

An H.264/MPEG-4 AVC encoder, using a single reference I-picture followed by predicted P-pictures, was used to code the SI-based PVS.[10] To achieve a specific compression ratio for the full II-frame, a pseudo bit rate R in $bits/s$ was defined for the SI-based PVS-encoding as

$$R = K \cdot L \cdot r \cdot f, \tag{5}$$

where f is the PVS pseudo frame rate and r the desired number of bits per pixel of the II-frame. See Table 2 for further details regarding the selection order evaluation.

4.1. SI-selection orders

It is possible to assess the different selections orders potential of enhancing the compression efficiency by evaluating the formed PVS using the cross-correlation coefficient c. This is based on the assumption that high correlation between consecutive PVS-pictures can efficiently be utilized to reduce redundancy by the motion compensation stage of the video

Table 1. Reference scene characteristics

Scene name	Degree of detail	Depth range	Fill factor
Apples	low	long	low
Twins	high	long	low
Car	low	short	high
Cuboid	high	long	high

Figure 6. Perspective projections of the defined reference scenes (a) Apples, (b) Twins (c) Car and (d) Cuboid.

Table 2. Setup for SI-selection order evaluation

Description	Value
II-frame resolution [pixels]	8192×4096
SI resolution - $K \times L$ [pixels]	512×256
EI resolution - $U \times V$ [pixels]	16×16
Bit rate range - r [bpp]	$[0.015, 1.5]$
Pseudo frame rate of PVS - f [fps]	25
Picture types in coded PVS	IPP...P

encoder. For this estimation, the cross-correlation coefficient c for two consecutive images $PVS(t)$ and $PVS(t-1)$ in the PVS was defined for $1 \leq t < U \cdot V$ as

$$c(t) = \frac{\sum\limits_{k=0}^{K-1} \sum\limits_{l=0}^{L-1} |X(t)| \cdot |X(t-1)|}{\sqrt{\left(\sum\limits_{k=0}^{K-1} \sum\limits_{l=0}^{L-1} |X(t)|^2\right)\left(\sum\limits_{k=0}^{K-1} \sum\limits_{l=0}^{L-1} |X(t-1)|^2\right)}}, \tag{6}$$

where

$$X(t) = PVS(t) - \overline{PVS(t)}$$

is the PVS-picture t after subtraction of its pixel mean, i.e. average brightness. The mean $\bar{c}$ and standard deviation σ_c was then calculated using

$$\bar{c} = \frac{1}{U \cdot V - 1} \sum_{t=1}^{U \cdot V - 1} c(t) \tag{7}$$

and

$$\sigma_c = \sqrt{\frac{1}{U \cdot V - 2} \sum_{t=1}^{U \cdot V - 1} (c(t) - \bar{c})^2}. \tag{8}$$

The cross-correlation only gives an assessment of the compression efficiency of the selection orders. To measure the actual quality, the color Peak Signal to Noise Ratio (PSNR) in dB was used according to

$$PSNR(II, \widehat{II}) = 20 \cdot log_{10} \left(\frac{255}{\sqrt{MSE}}\right), \tag{9}$$

where

$$MSE = \frac{1}{3 \cdot M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left|II(m,n) - \widehat{II}(m,n)\right|^2.$$

II and $\widehat{II}$ are the original and the compressed II-frame respectively.

4.2. SI- vs. EI-based PVS

From our previous work it is evident that the EI-based PVS coding scheme can not exploit the inherent redundancy of this II-frame structure.[6] This since the required header information becomes a substantial portion of the bit rate for coded low resolution PVS-pictures. However, as the number of EIs in the II-frame structure is reduced – and their resolution increases – the advantage of the SI-based PVS coding scheme was anticipated to decrease.[6] For an II-frame structure with a small set of high resolution EIs the EI-based PVS would outperform the SI-based. To evaluate this hypothesis a set of five II-frame structures was defined and named according to their combined EI- and SI- resolutions KxLxUxV: 512x256x16x16, 256x128x32x32, 128x64x64x64, 64x32x128x128, 32x16x256x256. Note that the II-frame resolution MxN, i.e. the KxLxUxV-product, was kept constant. Also, by using these II-frame structures a sufficient granularity was achieved for the cross-over point search while still keeping a moderate simulation time. Evaluation was performed using the difference in PSNR, i.e.

$$\Delta PSNR(r) = PSNR_{SI}(r) - PSNR_{EI}(r), \tag{10}$$

which was calculated for bit rates r where both of the two PVS approaches could be evaluated. See Section 5.1 for further discussions on why not all bit rates can be evaluated for all PVS approaches. Note that for an EI-based PVS, the PVS-picture resolution factor $K \cdot L$ in Eq. (5) is changed to $U \cdot V$.

4.3. Coding artifacts

All possible views are evaluated simultaneously when the calculations on the whole II-frame is performed. By this, a wide overview of the performance of the coding schemes is achieved. However, since the evaluated coding schemes all operate in different domains (EI, SI and II) it is also of importance to asses how the effects of the coding is perceived when the II-frame is seen through an II-display. Given the hypothesis that both PVS approach would be favored by a high PVS-picture resolution, an additional II-frame structure of 64x64x64x64 was defined to remove any advantages related to resolution. Thereby focus was placed solely on the effects of each coding scheme. From each of thee four reference scenes, and the corresponding II-frames, a set of five snapshot views was decoded for evaluation: front-, up-, down-, left- and right. Each snapshot view, or view image (VI), was constructed using one pixel per EI as illustrated in Fig. 1, i.e. with a resolution of $K \times L$ pixels, according to

$$VI_{\mathbf{E}}(k,l) = II(k \cdot U + \frac{U}{2} + \Delta u(k,l), l \cdot V + \frac{V}{2} + \Delta v(k,l)), \tag{11}$$

where $\mathbf{E} = [\mathbf{E}_x, \mathbf{E}_y, \mathbf{E}_z]^T$ is the 3D-point in viewing space from which the decoded II-frame was viewed. The II-display is located such that its pixel array containing the II-frame content coincides with the xy-plane. The pixel offsets of each lens, $\Delta u(k,l)$ and $\Delta v(k,l)$, were defined as

$$\Delta u(k,l) = d \cdot \frac{(\mathbf{l}_x(k,l) - \mathbf{E}_x)}{\mathbf{l}_z(k,l) - \mathbf{E}_z}$$

and

$$\Delta v(k,l) = d \cdot \frac{(\mathbf{l}_y(k,l) - \mathbf{E}_y)}{\mathbf{l}_z(k,l) - \mathbf{E}_z}$$

where d is the distance between the pixel and lens array in the II-display (see Fig. 1) and $\mathbf{l}_x(k,l)$, $\mathbf{l}_y(k,l)$ and $\mathbf{l}_z(k,l)$ are the 3D-space coordinate for the (k,l)-th lens center. A pinhole approximation was used for the lenses. For non-integer pixel offsets, linear interpolation was performed. The view images from Twins are presented in Fig. 7.

The five VIs were evaluated using the gray scale Mean Structural Similarity index (MSSIM),[11] which was selected for its good correspondence with subjective tests on image coding quality. Also, the fact that the MSSIM is motivated by a sliding window principle allows it to capture inter-pixel effects that the pixel-difference approach of the PSNR metric is unable to do. The five MSSIM values were later averaged into a single value for each bit rate r according to

$$\overline{MSSIM}(r) = \frac{1}{5} \cdot \sum_{n=0}^{4} MSSIM \left(VI_{\mathbf{E}_n}, \widehat{VI}_{\mathbf{E}_n}(r) \right) \tag{12}$$

where $VI_{\mathbf{E}_n}$ and $\widehat{VI}_{\mathbf{E}_n}(r)$ were the original and coded view image respectively. The coded view images from each coding scheme were also visually inspected to identify any characteristic pattern in coding artifacts.
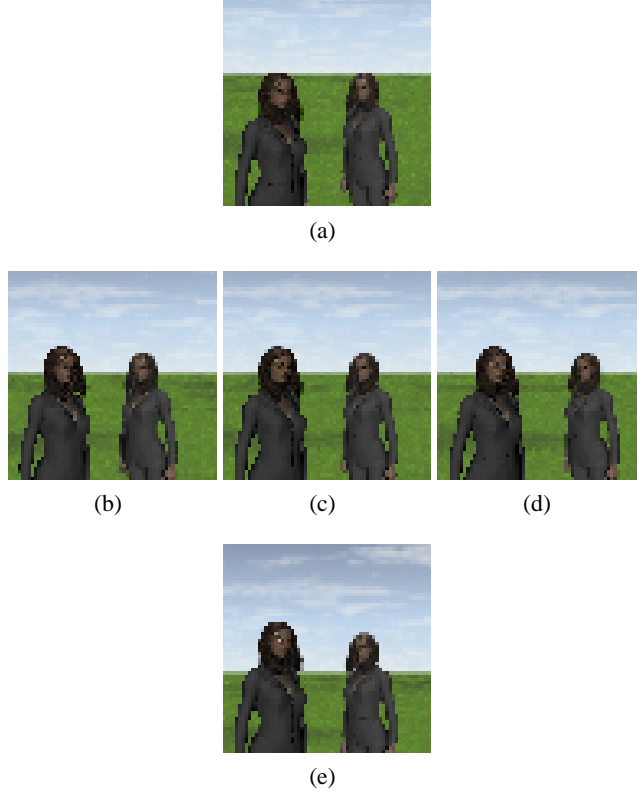
Figure 7. Reference II-frame Twins decoded from view point (a) $\mathbf{E}_0$=up, (b) $\mathbf{E}_1$=left, (c) $\mathbf{E}_2$=front, (d) $\mathbf{E}_3$=right and (e) $\mathbf{E}_4$=down.

Table 3. Mean and standard deviation ($\bar{c}$,σ_c) of cross-correlation coefficient c

	Apples		Twins		Car		Cuboid	
	$\bar{c}$	σ_c	$\bar{c}$	σ_c	$\bar{c}$	σ_c	$\bar{c}$	σ_c
Row	0.87	0.13	0.88	0.15	0.72	0.17	0.63	0.17
Column	0.78	0.21	0.86	0.16	0.37	0.39	0.44	0.19
Parallel	0.90	0.09	0.92	0.03	0.73	0.14	0.65	0.13
Zig-zag	0.79	0.15	0.88	0.08	0.48	0.31	0.38	0.19
Spiral	0.86	0.13	0.90	0.09	0.61	0.25	0.53	0.23

5. RESULTS

5.1. SI-selection orders

In Table 3, the mean $\bar{c}$ and standard deviation σ_c of the cross-correlation coefficient c is calculated using Eq. (7) and (8) and compiled for all combinations of selection orders and reference II-frames. Forming a PVS using the parallel selection order results in the highest average, and the lowest standard deviation, of c. This indicates, according to the assumption made, that forming a PVS using the parallel selection order should also be most beneficial for compression.

In Fig. 8, the objective quality in terms of PSNR for the different selection orders are shown. The dotted and dotted-square curves are included for reference and correspond to the JPEG 2000 coded II-frame and the EI-based PVS scheme[5] respectively.

When studying the average PSNR over the whole range of bit rates, the parallel selection order proves to be best with 39 dB, 34 dB, 33 dB and 22 dB for Apples, Twins, Car and Cuboid respectively. However, as seen in Fig. 8 the difference between the selection orders is small. Compared to parallel the row selection order is for example 0.25 dB, 0.14 dB,

0.34 dB and 0.02 dB worse for the four reference II-frames. For Apples and Car, there is significantly better performance of *horizontal* selection orders - row and parallel - over the rest; less so for Twins and Cuboid. This is believed to be an effect of the foremost *horizontal* object positioning in the used reference scenes. For scenes containing an extensively random object positioning, the difference between selection orders is anticipated to be even lower.

In some PSNR-curves there is a lack of data points for low bit rates r, e.g. in Fig. 8 (d). This is due to the large disparity introduced between neighboring SIs combined with the maximum allowed value of the quantization parameter QP in the H.264 standard. Scenes with high detail, depth and fill factor introduce a large disparity in the SIs which in turn cause large differences between consecutive PVS-pictures. If the difference is sufficiently large, the residual after motion compensation still contains high values. Given the maximum value of the H.264-quantization parameter QP the encoder is unable to reduce these values enough in the quantization stage. Thus, the encoder fails to provide the demanded bit rate r. If it would be possible to set a value of $QP > 51$, and thereby quantize the residual from the motion compensation stage harder, a lower bit rate than what is shown would be achievable. Of course this would come at the price of a lower PSNR.

In our previous work we have shown that our proposed method outperformed both JPEG 2000 and the EI-based PVS coding scheme for 0.15 bpp $\leq r \leq$ 1.5 bpp. The PSNR-curves in Fig. 8 indicate that this is also true for 0.015 bpp $\leq r <$ 0.15 bpp, regardless of selection order.

5.2. SI- vs. EI-based PVS

Fig. 9 shows the PSNR gain, calculated using Eq. (10), that the SI-based PVS approach have over the EI-based when the II-frame structure is varied. The results confirm our hypothesis that when the EI-resolution is increased the advantage of the SI-based PVS scheme is more and more reduced. The results also indicate that the cross-over point of PVS-scheme efficiency occurs when the EI and SI resolutions are approximately equal. For Twins though, the cross-over point occur practically at 128x64x64x64 as the $\Delta PSNR$ is low over the whole range of bit rates.

However, there is also another factor to consider when choosing PVS approach that is related to the content of the II-frame. The orthographic property of the SI is believed to distribute the coding artifacts of the SI-based PVS scheme evenly over the whole depth range. Thus, if there is high detail at long depths, as for Twins and Cuboid, this will compete for the bit rate on equal term with any detail at short depths. For the EI-based scheme this is not the case since even if there is high detail at long depths it is blurred out due to the averaging nature of the EI's linear perspective. Since the tested reference scenes where synthesized using a pinhole lens approximation, the II-frames have an infinite depth of field. In a physically II-camera on the other hand, distant objects would be blurred due to the finite depth of field thereby reducing the difference between the two PVS approaches.

When examining Fig. 9 it can be concluded that the main factor in selecting PVS-scheme relates to the II-frame structure. If $K \cdot L > U \cdot V$ the SI-based PVS should be used otherwise the EI-based PVS. That is, from a PSNR perspective the PVS approach that results in the largest resolution of the PVS-pictures should be preferred.

5.3. Coding artifacts

The two results presented in Section 5.1 and 5.2 were both based on the complete II-frame. To further examine the EI- and SI-based PVS approaches introduced coding artifacts, the II-frame structure 64x64x64x64 is studied. In Fig. 10, the $\overline{MSSIM}$ values calculated from Eq. (12) are presented for the two PVS approaches and the JPEG 2000 coding of the II-frame. The two PVS approaches converge to MSSIM = 1 as the bit rate r is increased, which is expected for any coding scheme for a sufficiently high bit rate. Note that the SI-based PVS scheme performs equal to the EI-based for Car that has a short depth range. Objects on a medium distance from the II-camera have similar projection size and translation speed in consecutive PVS-pictures for both PVS approaches. Consequently, the lack of objects at long distance makes the SI-based PVS scheme comparable to the EI-based PVS scheme's feature, i.e. prioritizing objects closer to the II-camera. For Cuboid and its repetitive character, the orthographic property of the SI results in a PVS that is coded more efficiently than the EI-based PVS scheme. This due to that a significant number of the SI-based PVS-pictures captures the sub-cubes in the cuboid aligned in the depth direction. Close sub-cubes occlude distant ones, which results in a large portion of the PVS-picture containing background and therefore being easy to code. The set of EI-based PVS-pictures on the other hand are much more similar in character to each other: each showing a slightly translated perspective of the complex cuboid. Thus, the EI-based PVS scheme applied to Cuboid results in a less efficient coding.
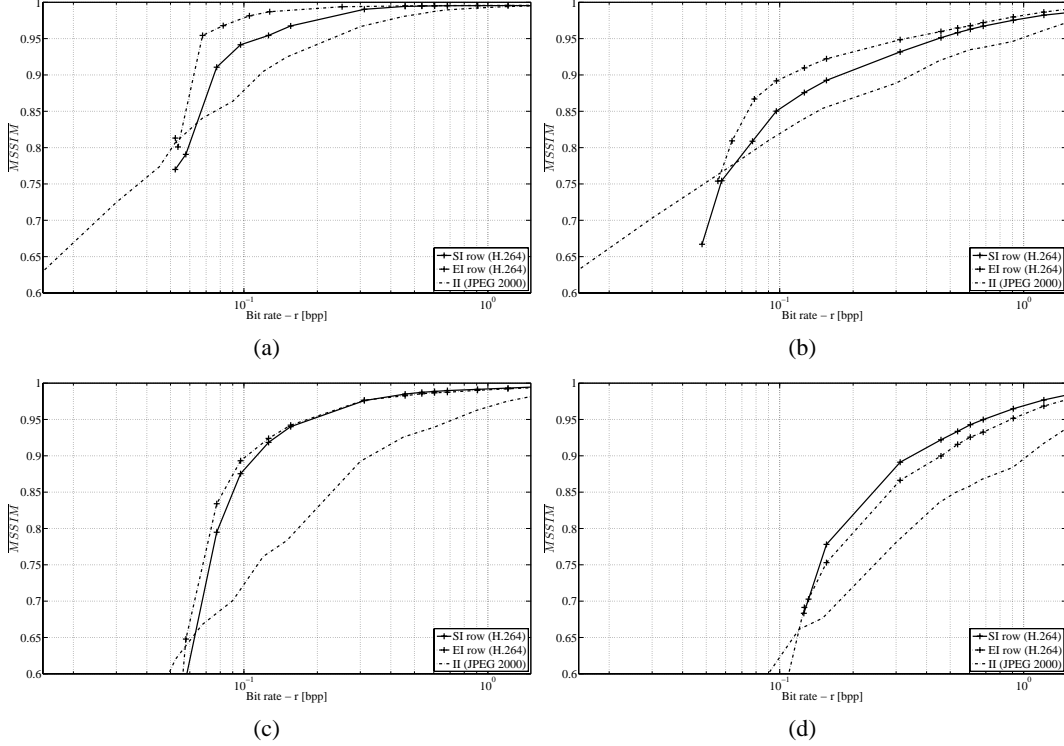
In Fig. 11 the specific artifact characteristics of the coding schemes are illustrated for a bit rate $r = 0.15$ bpp. As can
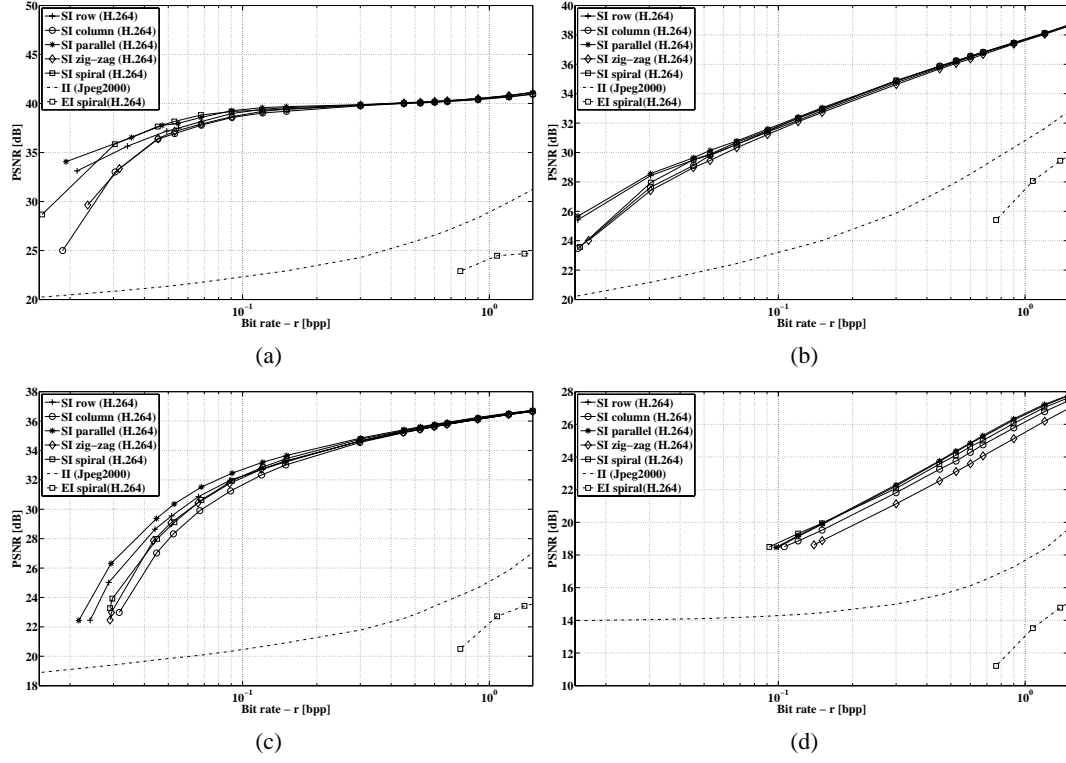
Figure 8. Selection order evaluation at II-frame structure 512x256x16x16 using PSNR, Eq. (9). (a) Apples, (b) Twins, (c) Car and (d) Cuboid.
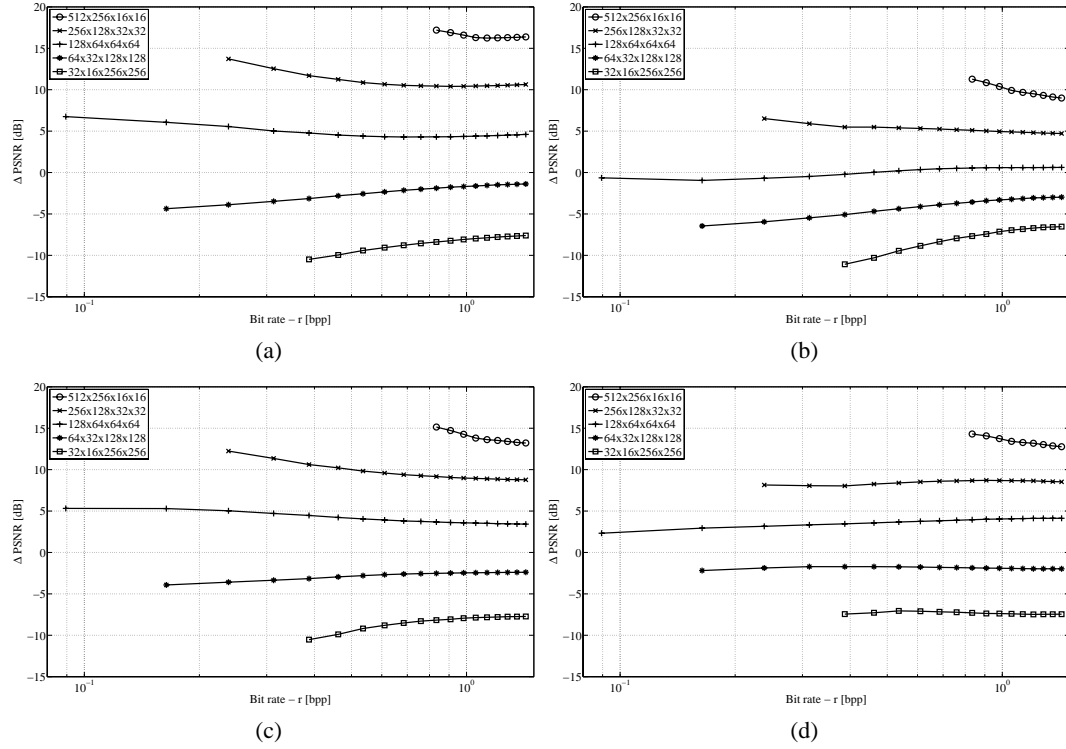
Figure 9. Comparison of SI- and EI-based PVS-scheme using $\Delta PNSR$, Eq. (10). (a) Apples, (b) Twins, (c) Car and (d) Cuboid.

Figure 10. $\overline{MSSIM}$ evaluation of II-frame structure 64x64x64x64, using Eq. (12). (a) Apples, (b) Twins, (c) Car and (d) Cuboid.

be seen in Fig. 11 (d), and as could be expected from examining Fig 8 and 10, JPEG 2000 coding of the II-frame results in the worst quality. This due to the lack of directional prediction in JPEG 2000, which the motion compensation stage in the two PVS approaches allow for. When examining each VI, the character of the coding artifacts somewhat resembles additive gaussian noise with a standard deviation that increases with increasing depth. The difficulty in coding distant objects comes from that these objects, due to the lens array, spread out over a larger portion of the II-frame than close objects. This makes the introduced redundancy, which spans EIs that are further apart, more difficult to address for JPEG 2000.

When studying the artifacts caused by the SI-based PVS coding scheme (with row selection order) in Fig. 11 (b), a blurring over the whole depth range can be seen. In Twins this is especially evident in the face of the woman closest to the II-camera. As could be expected, for images with less detail the blurring effect is also less pronounced. Apples and Car for example, contain objects with textures that already are relatively low frequency and thus show less coding artifacts. There is also a spread in colors around object edges, much like the mosquito noise that is well-known from many DCT-based coding schemes. The level of this noise tends to decrease with increase depth – compare the furthest and closest objects in Apples and Woman respectively.

In the EI-based PVS scheme (also with row selection order), the noise has slightly different characteristics as Fig. 11 (c) illustrates. The coding artifacts have a significant horizontal component, which is not the case for the mosquito-like artifacts introduced by the SI-based PVS. The level of the EI-based PVS noise also increase in level with *increasing* depth, as can be seen when comparing the edges of distant and close objects in Apples and Twins. In the complex Cuboid, the horizontal smearing combines with the penalizing of distant objects, which results in a coded image that looks unwashed compared to the SI-based version.

6. CONCLUSIONS

In this paper we have evaluated five selection orders for our previously proposed SI-based PVS coding scheme on a set of four full color II-frames. Our results show that the parallel selection order is preferred. They also indicate that regardless of

Figure 11. Characteristic coding artifacts at $r = 0.15$ bpp for (b) SI-based PVS, (c) EI-based PVS and (d) JPEG 2000 coded II-frame. (a) uncoded front views for reference.

used selection order, the proposed scheme allows for a coding at a bit rate $r \geq 0.07$ bpp – corresponding to a compression ratio of up to 342:1 – while still keeping the $PSNR$ above 30 dB for all but the most complex reference II-frame. A result that well exceeds what is possible to achieve with JPEG 2000 or a EI-based PVS coding scheme for the investigated II-frame structure.

We also studied the cross-over point between EI-based and SI-based PVS coding and conclude that the PVS approach which results in the highest resolution of the PVS-pictures should be preferred. Thereby minimizing the relative portion of the bit rate that must be set aside for mandatory meta data such as coded picture header information.

Differences in coding quality for the evaluated schemes was also studied and presented using MSSIM as well as visual inspection. The examination was performed using an II-frame structure which would eliminate any advantages that each PVS approach would have from a higher PVS-picture resolution. A significant difference in coding artifact character between the two PVS approaches was identified. The artifacts stemming from the EI-based PVS tended to increase as a function of increasing object depth. The SI-based PVS on the other hand distributed its coding artifacts more homogenously within the depth range. For some reference scenes it decreased slightly as a function of increased object depth. It should be noted that even though the different artifact characteristics are a consequence of the selected coding scheme, the way they manifest are to a small degree also a function of the chosen visualization method. Thus, further studies should be performed, using physical II-displays coupled with subjective tests, to give a complete view of how the coding schemes affect the perceived 3D-quality of the II-frame.

ACKNOWLEDGMENTS

REFERENCES

1. B. Javidi and F. Okano, eds., *Three-Dimensional Television, Video, and Display Technologies*, Springer, 2002.
2. M. C. Forman, N. Davies, and M. McCormick, "Continuous parallax in discrete pixelated integral 3d displays," *Optical Society of America* , 2002.
3. M. C. Forman, *Compression of Integral Three-Dimensional Television Pictures*. PhD thesis, De Montfort University, September 1999.
4. 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.
5. S. Yeom, A. Stern, and B. Javidi, "Compression of 3d color integral images," *Optics Express* **12**, pp. 1632 – 1642, April 2004.
6. 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*, IEEE, October 2006.
7. C. Wu, A. Aggoun, M. McCormick, and S. Kung, "Depth extraction from unidirectional integral image using a modified multi-baseline technique," *Stereoscopic Displays and Virtual Reality Systems IX* **4660**, pp. 135 – 145, May 2002.
8. J.-H. Park, S. Jung, H. Choi, Y. Kim, and B. Lee, "Depth extraction by use of a rectangular lens array and one-dimensional elemental image modification," *Applied Optics* **43**, pp. 4882– 4895, September 2004.
9. 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, June 2005.
10. "x264 - a free h264/avc encoder. core: 38 svn-341." http://developers.videolan.org/x264.html, December 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.