Roger Olsson, Mårten Sjöström, and Youzhi Xu. A combined pre-processing and H.264-compression scheme for 3D integral images. In *Proceedings of International Conference on Image Processing*, IEEE, pp. 513 -- 516, Atlanta (GA), USA, October, 2006.

http://dx.doi.org/10.1109/ICIP.2006.312389

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

The next evolutionary step in enhancing video communication fidelity is taken by adding scene depth. 3D video using integral imaging (II) is widely considered as the technique able to take this step. However, an increase in spatial resolution of several orders of magnitude from todays 2D video is required to provide a sufficient depth fidelity, which includes motion parallax. In this paper we propose a pre-processing and compression scheme that aims to enhance the compression efficiency of integral images. We first transform a still integral image into a pseudo video sequence consisting of sub-images, which is then compressed using an H.264 video encoder. The improvement in compression efficiency of using this scheme is evaluated and presented. An average PSNR increase of 5.7 dB or more, compared to JPEG 2000, is observed on a set of reference images.

***Index Terms*—** Three-dimensional displays, Image coding

1. INTRODUCTION

Three-dimensional (3D) video has for decades been pursued as the future of video communication. Integral imaging (II) is an autostereoscopic 3D technique that places no requirements of special glasses on the viewer at the expense of a large increase in spatial resolution. However, this increase give rise to a large amount of redundancy, which opens up for potentially efficient compression.

Other techniques for providing the sensation of perceived depth have also been presented [1]. Recent years progress in the field of liquid crystal display (LCD) research has transformed II into a promising implementation path for 3D video. A few II-based prototypes have also been presented [2, 3, 4].

Conceptually, an II-camera differs from a two-dimensional camera in that the smallest display unit is not a single RGB-pixel but a lens coupled with a set of RGB-pixels; thereby increasing the spatial resolution requirement. Together the lenses act as an optical multiplexer, implicitly storing scene depth by redirecting light from different view angles to different pixel subsets. The pixels covered by any given lens captures a low resolution projection of the scene, a so-called elementary image (EI). Combined the EIs form the complete II-frame. An II-display inverse this process by a similar lens- and pixel array set-up: different views are demultiplexed and distributed into different directions of the viewing space, as illustrated in Fig. 1. Given that two different views are seen by the left and right eye respectively binocular parallax is achieved, which is required for perceiving scene depth. To achieve a closer approximation to viewing a real 3D scene a larger number of pixels per lens than two is required. This allows the user to see different parts of the captured scene when changing viewing position relative to the display, so-called motion parallax.

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

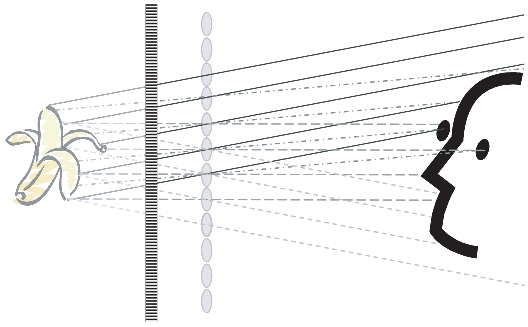


Fig. 1. II-display with pixel and lens array. Rays passing a given lens corresponds to an elementary image (EI). Parallel rays from different lenses corresponds to a sub-image (SI).

For a system providing only binocular parallax, the increase in the number of pixels compared to 2D is only twofold. However, Forman et al. [5] have shown that in order for the II's inherently discrete views to be perceived as continuous motion parallax, an increase of more than ten times in both the vertical and horizontal direction is required. Keeping the number of pixels per II-frame constant, a trade-off between depth and spatial resolution can be made.

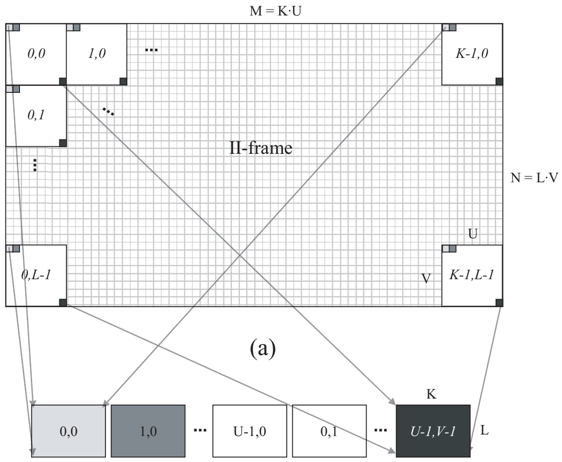
The increased data requirement of II have been addressed using different approaches in previous works. Forman [6] focused on compressing horizontal parallax only (HPO) II-frames by quantizing and entropy coding the coefficients from a 3D-DCT. Yeom et al. [7], on the other hand, pre-processed II-frames with a low number of high resolution EI by extracting all EIs according to a pre-defined order. These were stacked together, forming a motion picture, and encoded using an MPEG-2 video encoder.

In this paper we propose a new scheme that is adapted to full parallax II-frames with a large number of low resolutions EIs. The scheme transforms an II-frame into a set of sub-images that when fed to an H.264 (also known as MPEG-4 AVC) video encoder results in a significantly increased compression efficiency compared to other schemes. The structure of the II-frame is described in Section 2. In Section 3 the proposed scheme is then presented. This is followed by the experimental setup used to evaluate the scheme, i.e. chosen II-frame structure, used reference II-frames, selected coding parameters etc., in Section 4. The resulting coding efficiency is presented and concluding remarks are given in in Section 5 and Section 6 respectively.

2. II-FRAME STRUCTURE

An II-frame with a resolution of $M \times N$ pixels is defined as

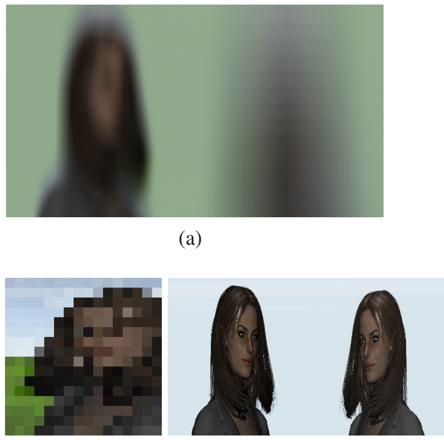
$$II\left(m,n\right), \tag{1}$$



(a)

(b)

Fig. 2. II-frame structure showing (a) the II-frame and the EIs, and (b) the forming of SIs.



(a)

(b) (c)

Fig. 3. Example content showing (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.

where $m = 0, 1, ..., M-1$ and $n = 0, 1, ..., N-1$ are the horizontal and vertical positions of an II-frame pixel respectively. Thus, $II(m_0, n_0) = [ii_R, ii_G, ii_B]^T$ corresponds to the RGB-color of the m_0-th column and n_0-th row pixel.

A number of EIs constitutes an II-frame, as illustrated in Fig. 2 (a). The shape and position pattern of the EIs might differ between II-techniques. A HPO II-systems might use EIs that are rectangular, strip-like and positioned adjacent to each other only in the horizontal direction [6]. Full parallax II-systems on the other hand can use both circular EIs positioned in a hexagonal pattern as well as rectangular EIs positioned in a rectangular pattern [2, 8].

In this paper, we use rectangular EIs positioned in a rectangular pattern, i.e. an 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 $K \times L$ EIs has a resolution of $U \times V$ pixels.

Within the field of depth extraction, II-techniques have been used because they allow for using one single camera instead of several, thereby removing the multiple camera calibration stage [9, 10]. To enhance the depth estimation, the concept of sub-images was introduced, which extracts and combines II-frame pixels sharing the same relative horizontal offset to the EI centers. Extending this concept to also include relative vertical offset is straightforward and is in this paper referred to as a complete sub-image (SI) 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. Thus, each of the $U \times V$ SIs has a resolution of $K \times L$ pixels. In Fig. 1, pixels that are intersected by parallel lines of the same line-style belong to the same SI. In Fig. 2, a specific SI is illustrated using a specific gray level.

Figure 3 presents an example of II-structure content, corresponding to the scene Twins shown in Fig. 4 (a). The full II-frame is shown in Fig. 3 (a) whereas Fig. 3 (b) presents one of the EIs, containing a low resolution *perspective* projection of the scene. Figure 3 (c) depicts an SI that, contrary to an EI, contains a high resolution *orthographic* projection of the scene. I.e. in an SI, an increased object depth does not result in an increased size of the object's projection.

3. PROPOSED SCHEME

Utilizing video coding tools was shown to be efficient when applied to a small set of high resolution EIs [7]. However, for II-techniques that use a large set of low resolution EIs this is not suitable because each compressed image requires meta data to enable a proper decode. For low resolution images this overhead becomes a substantial part of the bit rate, thus penalizing image quality.

In our proposed scheme we apply a pre-processing operation that transforms the II-frame into an SI-based pseudo video sequence (PVS), i.e. the set of SIs is used to create a sequence that is a function of space and angle, not of time. Compared to using an EI-based PVS, an improved coding efficiency is achieved since the meta data portion for higher resolution SIs are proportionally lower. Compared to still image coding, a larger part of the II-frame's inherent redundancy between views is exposed. To reduce this redundancy we apply an H.264 video encoder to the PVS, thereby addressing the spatial redundancy of the II-frame using tools aimed at reducing both spatial as well as temporal redundancy. I.e. when applied on a SI-based PVS, the motion compensation stage of H.264 performs in fact a disparity compensation. Due to the orthographic property of the SI, II-frames stemming from scene with long depth transforms into large disparity between consecutive PVS-images.

4. EXPERIMENTAL SETUP

To evaluate the proposed scheme, four reference scenes were defined with different degrees of detail, depth and fill factor. Twins - high detail, long depth and low fill factor; Apples - low detail, long depth and low fill factor; Car - low detail, short depth and high fill factor; Cuboid - high detail, long depth and high fill factor. From these scenes, four II-frames were synthesized using 24 bits per pixel [11]. Perspective projections of these scenes are shown in Fig. 4.

Three other schemes were used as a basis of comparison: an

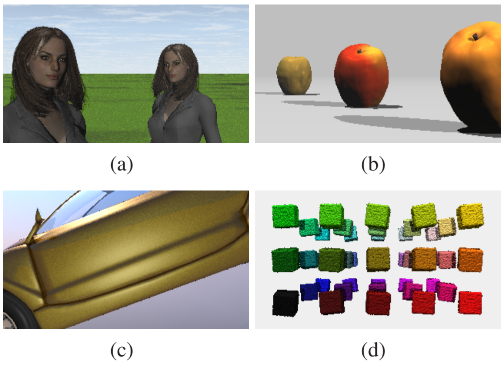


Fig. 4. Perspective projections of the defined reference scenes (a) Twins, (b) Apples, (c) Car and (d) Cuboid.

Table 1. Experiment setup

II-frame resolution - $M \times N$ [pixels]	8192×4096
Bit rate range [bpp]	$[0.015, 1.5]$
Number of EIs - $K \times L$	512×256
EI resolution - $U \times V$ [pixels]	16×16
Pseudo frame rate of PVS [fps]	25

EI-based PVS coded using H.264 and the II-frame (without pre-processing) coded using JPEG and JPEG 2000 [13, 14].

A pinhole lens approximation was used to model the II-camera. Normally distributed noise $N(\mu = 0, \sigma = 1)$ was added to the red, green and blue component of each 24 bit pixel to emulate thermal noise in the capturing pixel array.

The PVSs was compressed using an H.264-encoder with only one reference I-frame followed by predicted P-frames [12]. The rest of the coding parameters were kept at default values. To achieve a specific compression ratio for the full II-frame, a pseudo bit rate R in $bits/s$ was defined for the PVS encoding as

$$R = M' \cdot N' \cdot bpp \cdot fps, \quad (4)$$

where $M' \cdot N'$ is the number of pixels in each PVS-image, fps the PVS pseudo frame rate and bpp the desired bits per pixel of the II-frame. For the SI-based PVS, $M' \cdot N' = K \cdot L$.

To evaluate the quality loss introduced by compression, the full color Peak Signal to Noise Ratio (PSNR), in dB, was used:

$$PSNR(II, \hat{II}) = 20 \cdot log_{10} \left(\frac{255}{\sqrt{MSE}} \right), \quad (5)$$

where

$$MSE = \frac{1}{3 \cdot M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left| II(m,n) - \hat{II}(m,n) \right|^2 \quad (6)$$

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

Table 1 contains further details about the experimental setup. Note that the trade-off between spatial and depth resolution was set such that the number of EIs approach the number of pixels of standard definition TV and the EI resolution approach the requirements of continuous parallax [5].

5. RESULTS

Figure 5 shows the PSNR of the proposed SI-based PVS scheme, together with the reference schemes. The proposed SI-based PVS scheme performs better over the greater part of the tested range of bit rates. For example, when comparing with JPEG 2000 over bit rates $[0.1, 1.5]$ bpp, an average PSNR increase of 7.8 dB, 13.7 dB, 11.9 dB and 5.7 dB can be observed for the four reference II-frames in Fig. 5 (a)-(d), respectively.

For complex scenes with high detail, depth and fill factor, the introduced disparity between neighboring PVS-images is high. This strains the motion compensation such that even if the H.264 quantization parameter is set to its maximum value ($QP = 51$), there still remains significant DCT-coefficients in the residual images. In Fig. 5 (b)-(d) this appears as incomplete curves for the SI-based PVS scheme, where a lower bit rate is impossible to achieve unless a nonstandard quantization parameter ($QP > 51$) is used.

Contrary to the SI-based PVS, the incomplete curves of EI-based PVS and JPEG coded II-frames are not a result of scene properties but a consequence of the minimum bit rate possible for the two schemes. The two vertical lines in Fig. 5 shows the bit rate achieved when coding a completely black II-frame, which is approximately the bit rate required for the meta data alone. I.e. a lower bit rate is not possible to achieve regardless of II-frame content. This clearly illustrates why the use of EI-based PVS and II-frames coded with JPEG is not beneficial for this type of II-structures. The minimum bit rates for SI-based PVS and JPEG 2000 coded II-frames is located *below* the range of bit rates shown in the figure.

Cuboid has the least, but yet significant, PSNR improvement of the four scenes when comparing the SI-based PVS with the reference schemes. This is due to that, for this scene, there exists less redundancy between views that can be exposed and reduced. A property caused by the scene's high detail and long depth.

6. CONCLUSION

In this paper we have presented a pre-processing and compression scheme that aims to enhance the compression efficiency of full color integral images with a large number of low resolution EIs. We have shown that the proposed scheme outperforms both still image coding, as well as other previously proposed coding schemes with improvements in average PSNR of more than 5.7 dB, when evaluated on a set of reference II-frames.

7. ACKNOWLEDGEMENT

The authors would like to acknowledge Jaime Vives Piqueres for modeling the 3D meshes used for defining the reference scenes Apples and Car.

References

[1] Editors: Bahram Javidi and Fumio Okano. *Three-Dimensional Television, Video, and Display Technologies*. Springer, 2002.

[2] Jun Arai, Fumio Okano, Haruo Hoshino, and Ichiro Yuyama. Gradient-index lens-array method based on real-time integral photography for three-dimensional images. *Applied Optics*, 37 (11), April 1998.

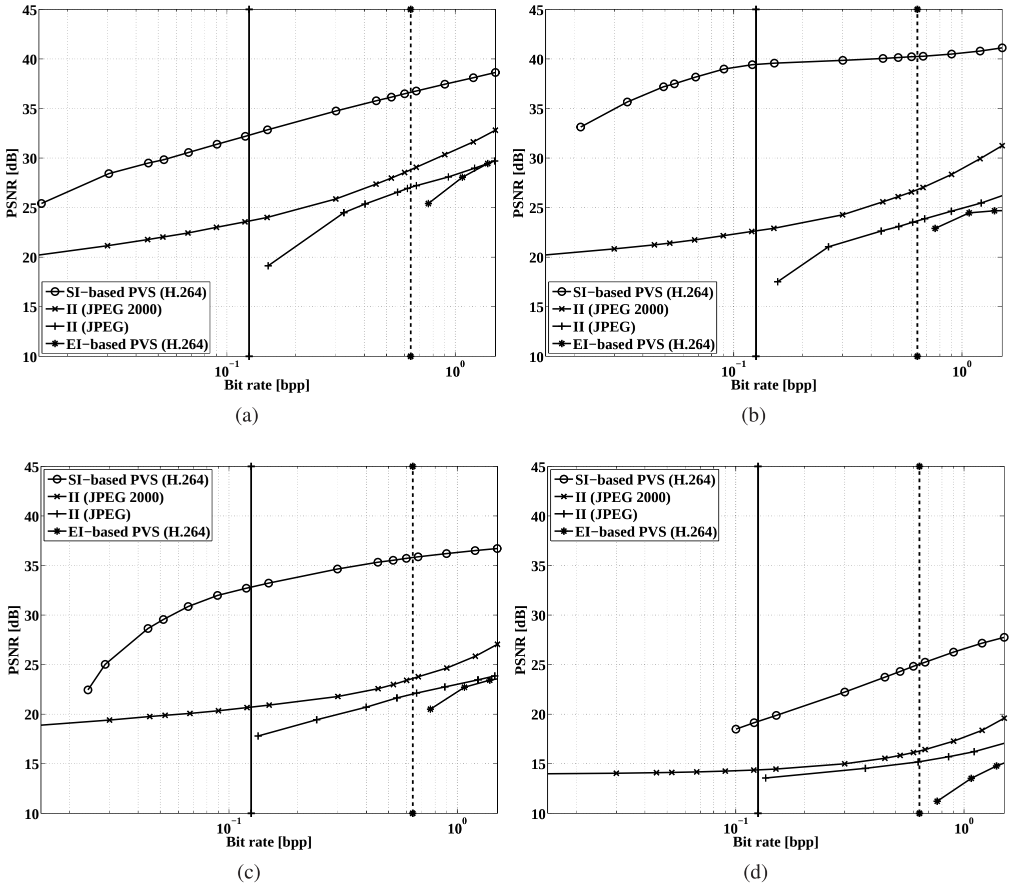


Fig. 5. Objective quality for reference II-frames (a) Twins, (b) Apples, (c) Car and (d) Cuboid. The two vertical lines indicate, from left to right, the minimum achievable bit rate for II (JPEG) and EI-based PVS (H.264).

[3] Hongen Liao, Daisuke Tamura, Makoto Iwahara, Nobuhiko Hata, and Takeyoshi Dohi. High quality autostereoscopic surgical display using anti-aliased integral videography imaging. In C. Barillot and D. Haynor, editors, *Proc. MICCAI 2004, LNCS*, pages 462–469. Springer-Verlag Berlin Heidelberg, 2004.

[4] Fumio Okano, Jun Arai, Kohji Mitani, and Makoto Okui. Real-time integral imaging based on extremely high resolution video system. In *Proceedings of the IEEE,*, volume 94. IEEE, March 2006.

[5] Matthew C. Forman, Neil Davies, and Malcolm McCormick. Continuous parallax in discrete pixelated integral 3d displays. *Optical Society of America*, 2002.

[6] Matthew C. Forman. *Compression of Integral Three-Dimensional Television Pictures*. PhD thesis, De Montfort University, September 1999.

[7] Sekwon Yeom, Adrian Stern, and Bahram Javidi. Compression of 3d color integral images. *Optics Express*, 12(8):1632–1642, April 2004.

[8] Jae-Hyeung Park, Sung-Wook Min, Sungyong Jung, and Byoungho Lee. Analysis of viewing parameters for two display methods based on integral photography. *Applied Optics*, 40 (29):5217–5232, October 2001.

[9] ChunHong Wu, Amar Aggoun, Malcolm McCormick, and S.Y. Kung. Depth extraction from unidirectional integral image using a modified multi-baseline technique. *Stereoscopic Displays and Virtual Reality Systems IX*, 4660:135–145, May 2002.

[10] Jae-Hyeung Park, Sungyong Jung, Heejin Choi, Yunhee Kim, and Byoungho Lee. Depth extraction by use of a rectangular lens array and one-dimensional elemental image modification. *Applied Optics*, 43(25):4882–4895, September 2004.

[11] Roger Olsson and Youzhi Xu. A ray-tracing based interactive simulation environment for generating integral imaging video sequences. In *Proceedings of Optics East 2005 - ITCOM - Three-Dimensional TV, Video, and Display IV*, volume 6016, October 2005.

[12] x264 - a free h264/avc encoder. core: 38 svn-341. http://developers.videolan.org/x264.html, December 2005.

[13] Independent jpeg group's free library for jpeg image compression. http://www.ijg.org/, December 2005.

[14] Michael D. Adams. The jasper transcoder. version 1.701.0. http://www.ece.uvic.ca/ mdadams/jasper/, December 2005.