
http://dx.doi.org/10.1109/3DTV.2008.4547821

© 2008 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.
EMPIRICAL RATE-DISTORTION ANALYSIS OF JPEG 2000 3D AND H.264/AVC CODED INTEGRAL IMAGING BASED 3D-IMAGES

Roger Olsson

Dep. of Information Technology and Media, Mid Sweden University, Sundsvall, Sweden

ABSTRACT

Novel camera systems producing 3D-images containing light direction in addition to light intensity is emerging. Integral imaging (II) is a technique on which many of these systems rely. The pictures produced by these cameras (II-pictures) are space-requiring in terms of data storage compared to their 2D counterparts. This paper investigates how coding the II-pictures using H.264/AVC and JPEG 2000 Part 10 (JP3D) affect the images in terms of rate-distortion as well as introduced coding artifacts. A set of four reference images are coded using a number of pre-processing and encoding variants, so called coding schemes. For low bitrates (<0.5bpp) the H.264/AVC-based coding schemes have higher coding efficiency, which asymptotically level of at higher bitrates in favor of JP3D. The JP3D coded 3D-images show less spread in quality than H.264/AVC, when quality is evaluated using PSNR as a function of viewing angle. However, the distortion induced by H.264/AVC is primarily localized to object boarders within the 3D-image, which in initial tests appear less visible than the JP3D coding artifacts that spread out evenly over the image. Extensive subjective tests will be performed in future work to further support the presented results.

I. INTRODUCTION

Prototypes of different types of 3D-camera systems have recently been presented [1]–[4]. These 3D-cameras – contrary to conventional 2D-cameras – capture not only light intensity but also light direction.

Integral imaging (II) is a 3D-imaging technique that relies on a lens array to separate intensity and direction of the captured light. The incoming light is relayed through the array within the II-camera and is finally captured by a high resolution pixel sensor. The combination of small lenses – all with slightly translated lateral positions – act as an optical multiplexer that implicitly stores the captured scene’s depth. The subset of pixels capturing light projected from a single lens is denoted elementary image (EI). The large number of low resolutions EIs combines to form the complete 3D-image captured by the pixel sensor. This type of 3D-image is hereafter denoted II-picture. Figure 1 presents a synthesized II-picture and an EIs is shown Figure 2 (a). The inverse process of the II-camera is performed by a II-display, which contains a lens array with similar properties. All views stored within the II-picture are distributed into their respective directions of the viewing space. Thus making II an auto-stereoscopic 3D technique with no requirements of special glasses for the viewer to perceive the 3D-image depth [5]. However, the cost is an increased pixel sensor resolution to accommodate the new directional information without loosing detail in the spatial information. Hence, coding becomes a beneficial if not required operation for many applications using II-pictures.

In this paper we will conduct an empirical rate-distortion analysis on two fundamentally different state-of-the-art coding standards applied to II-pictures: H.264/AVC and JPEG 2000 Part 10 (JP3D). The distortion induced on a set of II-pictures will be visually inspected as well as objectively measured. Stereoscopic view pairs will be presented to illustrate the type of artifacts that arise.

The paper is organized as follows. Section II presents the coding schemes, i.e. how the II-picture is pre-processed in order to fit as input to an H.264/AVC and JP3D encoder, constructed for volumetric images and video sequences respectively. In Section III, the II-pictures are presented together with the quality metrics with which they are evaluated. The evaluation results are presented in IV followed by a conclusion, which ends the paper.

II. CODING SCHEMES

Defining the II-picture and its components is a prerequisite for the subsequent presentation of all coding schemes. The
color II-picture is defined as:

\[ II(m,n), \]

where \( m = 0, 1, ..., M - 1 \) and \( n = 0, 1, ..., N - 1 \) are the horizontal and vertical positions of a II-picture pixel respectively. Hence, \( II(m_0, n_0) = [i_{IR}, i_{IG}, i_{IB}]^T \) corresponds to the RGB-color of the \( m_0 \)-th column and \( n_0 \)-th row pixel. Furthermore, the II-pictures used have quadratic EI-pictures. I.e. the EI at row \( k \) and column \( l \) is defined as

\[ EI_{k,l}(u,v) = II(k \cdot U + u, l \cdot V + v), \]

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. Complete sub-images (SI) are formed as subsets of pixels in the II-picture sharing the same relative horizontal and vertical offset to each EI center. An SI is thus defined as

\[ SI_{u,v}(k,l) = II(k \cdot U + u, l \cdot V + v) \]

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 A ray-space image (RI) is produced according to

\[ RI_{u,v}(k,l) = II(k \cdot U + u, l \cdot V + v) \]

The RIs belong to the ray-space [5], or ray-phase space [6], which has been proposed for different applications e.g. free-viewpoint television [7]. Figure 2 exemplifies EI, SI and RI selected from the II-picture Apples.

Each coding scheme is composed of two operations

1) a pre-processing step, which transforms the II-picture to a suitable format
2) an encoder adhering to a specific coding standard.

Forming the input for H.264/AVC and JP3D from an II-picture is handled by the same operation: exhaustively selecting all EIs, SIs or RIs. The selected images are either thought of as slices in a volumetric stack or pictures in a video sequence. High correlation between neighboring images is desirable to aid the redundancy reduction in the subsequent encoding. Hence, the order of selection is an important parameter for each coding scheme.

Fig. 2. Examples images extracted from II-picture Apples. (a) EI, (b) SI with its characteristic orthographic projection and (c) RI where the line segments slant angles correspond to their respective object’s depth.

Fig. 3. Two dimensional perspective projections of the four II-pictures Twins, Apples, Car and Cuboid.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lenslets (EI) - ( K \times L )</td>
<td>( 64 \times 64 )</td>
</tr>
<tr>
<td>El resolution - ( U \times V ) [pixels]</td>
<td>( 64 \times 64 )</td>
</tr>
<tr>
<td>Pixel sensor resolution - ( M \times N ) [pixels]</td>
<td>( 4096 \times 4096 )</td>
</tr>
<tr>
<td>Pixel sensor size [( mm^2 )]</td>
<td>( 25 \times 25 )</td>
</tr>
<tr>
<td>Pixel pitch - ( \delta ) [( \mu m )]</td>
<td>6.1</td>
</tr>
</tbody>
</table>

### III. EVALUATION SETUP

#### III-A. II-pictures

Four II-pictures (Twins, Apples, Car and Cuboid) was rendered using a II-camera model with pinhole lenslet approximation and normally distributed additive monochromatic noise \( N(\mu = 0, \sigma = 1) \). Perspective projections of the images are shown in Figure 3. The II-camera model’s pixel sensor was to approximate the CMOS 16MP sensor used in Canon EOS-1Ds Mark II. For further details about the II-camera and II-picture properties, see Table I. The model uses \( K = L = U = V = 64 \) making coding efficiency solely a function of image format (EI, SI or RI). The size of the produced set – and the resolution of each image – is constant regardless of selected image format. Hence, the portion of the bitrate required for bitstream metadata is also constant, which cancels out the II-picture structure’s affect on coding efficiency.

#### III-B. Evaluation methods

Two objective evaluation methods assess the distortion induced by the coding schemes: global PSNR and angle-dependent PSNR.

The single most adopted method to assess quality of II-pictures, and other 3D-image formats, is to globally apply the Peak Signal to Noise Ratio (PSNR) metric to the complete data set of pixels [8]–[11]. The advantage is that single scalar value are produced that aggregates the quality of the whole II-picture; lending itself well to comparative studies. However, aggregation is a weakness to gain detailed insight
into the distribution of introduced distortion. It is impossible to discern local properties within an II-picture from a single scalar value.

A more informative evaluation result is achieved by applying the PSNR in a viewing-angle-dependent manner \[12\]. Each measurement instance produces a quality matrix where each element represents the quality of the II-picture perceived by a virtual viewer from a specific viewing angle at an infinite viewing distance from a virtual 3D-display. That is, each SI pair – from both the original and coded II-picture – is used as input to the PSNR.

In addition, visual examination of the coding artifacts characteristics is also performed. Initially, stereoscopic image pairs synthesized from the II-pictures presented for cross-eyed viewing are used. This rudimentary autostereoscopic technique allows for inter alia evaluating the coding-induced distortion’s depth distribution. More extensive subjective tests will be performed in future work using state-of-the-art autostereoscopic displays.

III-C. Coding parametrization

The H.264/AVC and JP3D encoding was performed using x264 and the JPEG 2000 Part 10 reference software \[13\], \[14\] in the bitrate range \( r = \{0.015, \ldots, 1.5\} \) bpp. Note that the 3D-image is transformed into a video sequence and hence properties such as framerate, GOP-structure, motion vector search area etc. are all important coding parameters; default values for these were 25fps, IPPP... and \( 16 \times 16 \) pixels. The coding parameters for JP3D was parameterized as follows. A 5-level decomposition in all three dimensions was performed using the 5x3 wavelet kernel. Maximum tilesize was selected to maximize coding efficiency. Coding using JPEG 2000 was also performed for reference using similar coding parameters as JP3D. the II-pictures was coded without any pre-processing operations by the Kakadu software framework \[15\].

IV. RESULTS

Figure 4 acts as a general benchmark for the evaluated coding schemes. The two extremes in terms of 3D-image complexity (Apples and Cuboid) are presented. Additional graphs and stereoscopic views can be found at the paper’s accompanying web page \[^1\]. The global PSNR graphs show the H.264/AVC-based coding performing better for low bitrates (<0.5 bpp) with as much as 2dB. However, at higher bitrates the group of H.264/AVC-based coding schemes is restrained from producing higher PSNR. The asymptotic characteristic is caused by low complex 3D-images and the structure of H.264/AVC encoding; even when using the minimum quantization parameter (QP=1) the encoder still introduces distortion due to the fixed quantization matrix. However, for high complex 3D-images (such as Cuboid) this problem manifests at higher bitrates, which allows H.264/AVC to keep its advantage relative to JP3D even for medium bitrates (0.5 bpp – 1.5 bpp).

A quick study of the global PSNR for Apples at 0.5 bpp could lead to the conclusion that the two EI-based coding schemes (regardless of selected coding standard) would give comparable coding artifacts. However, this is jumping to conclusions as evident by studying the stereoscopic view pairs of Figure 5 – both extracted from the II-picture Apples coded at 0.5 bpp. Figure 5 (b) shows much more prominent distortion than subfigure (a), which contains artifacts that are hardly visible. Hence, the global PSNR is not sufficient to describe the coding schemes’ effects on II-pictures.

The angle dependent PSNR provides additional information about the characteristics of the coding schemes. Applying the metric on the two coded II-pictures being the source of the images in Figure 5 produce a quality matrix with \( 64 \times 64 = 4096 \) PSNR values as elements. Figure 6 shows the minimum and maximum PSNR values from all angles (or SIs) at each tested bitrate. The JP3D-based coding scheme gives a narrower range of PSNR-values compared the H.264/AVC scheme, which has higher maximum and lower minimum values. A closer study of the schemes from each view angle gives at hand that the distortion produced
two approaches H.264/AVC produces less visible distortion by localizing distortion to object borders. Future work will include further studies on optimizing coding parameters and an extensive subjective test using autostereoscopic displays to visualize the coded 3D-images aiming to support these initial findings on perceived coding induced distortion.

ACKNOWLEDGEMENT

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

VI. REFERENCES