ACQUISITION, PROCESSING AND CODING OF 3D HOLOSCOPIC CONTENT FOR IMMERSIVE VIDEO SYSTEMS

Amar Agooun1, Obaidullah Abdul Fatah2, Juan C Fernandez2, Caroline Conti3,4, Paulo Nunes3,4, Luis Ducla Soares3,4

1University Bedfordshire, Luton, UK
2Brunel University, London, UK
3Instituto de Telecomunicações, 4University Institute of Lisbon – ISCTE-IUL, Lisbon, Portugal

ABSTRACT

In the continuous effort to develop new multimedia content formats able to support more immersive and close to reality user experiences, 3D holographic imaging became lately an appealing technology, allowing a more natural and immersive 3D sensation with continuous full motion parallax, opening also new post-processing degrees of freedom, such as refocusing or viewing angle adjustments after content acquisition. In this context, this paper presents some novel achievements in 3D holographic content acquisition and post-processing. An improved coding solution for 3D holographic video based on the High Efficiency Video Coding (HEVC) standard and the recently proposed self-similarity compensated prediction concept is also presented that substantially improves the coding performance of HEVC for this type of content for both Intra- and Inter-coding.

Index Terms — 3D holographic video, integral imaging, 3D holographic video coding, HEVC, spatial prediction, self-similarity compensated prediction

1. INTRODUCTION

Three-dimensional (3D) imaging systems are attractive for the scientific community, entertainment and display industry to open a new market [1]. 3D images can be applied in broadcasting, communications and many other areas [2]. There are many technologies developed for 3D imaging systems. 3D holographic imaging (also referred to as Integral Imaging) as a spatial imaging method is a strong candidate for next generation 3D visualization systems [3].

3D holographic/plenoptic/Integral imaging systems find their origin from Integral Photography which was first proposed by Lippmann [4] in 1908 as a very promising method for capturing and reproducing three-dimensional images [5]-[7]. This technique uses the principle of “Fly’s eye” and hence allows natural viewing of objects. Unlike stereo imaging [8][9], 3D holographic imaging creates physical duplicates of the light field, being a true 3D technique. Compared with Holographic Imaging [10], it uses incoherent radiation and forms an image that is a sampled representation of the original object space and in full color. A flat panel display, for example one using LCD technology, can be used to reproduce the captured intensity modulated image and a microlens array (MLA) re-integrates the captured rays to replay the original scene in full color and with continuous parallax in all directions. The 3D content can be viewed by more than one person and independently of the viewer’s position.

In this paper, some novel achievements in the fields of acquisition, processing and coding of the 3D holographic video are presented.

2. 3D HOLOSCOPIC CONTENT ACQUISITION

3D holographic imaging is based on concurrent capture of many different views of a 3D scene by using a MLA, as shown in Figure 1(a). Under each micro-lens, there are certain pixels to depict different viewpoints. In order to achieve depth control, an objective lens or imaging lens is placed before the MLA and sensor. The MLA can either be fabricated onto the sensor or be placed in front of the relay system [12][13]. The benefits of using a relay system include: i) interchangeable MLA, ii) no need to access to sensor, and iii) easy alignment. However, the nonlinear distortion of the camera based on a relay system might hinder its applications. Both intrinsic and extrinsic factors in the integrated camera are contributing for the distortion. First, the lens radial distortion of the lens specific feature, which is the main intrinsic factor in the system; this will occur in the two main ways, which are barrel and pincushion effects. Second, is the MLA perspective distortion, which happens from the mounting procedure (extrinsic) of the MLA, which can induce perspective errors. The nonlinear distortion is critical for the 3D scene reconstruction and must be corrected before processing. In this paper the distortions are corrected using a third-party program.

Figure 1. (a) Layout of the single aperture 3D Integral Imaging system, (b) 3D Integral Imaging camera. PL: Prime lens, MLA: microlens array, RL: relay lens

978-1-4799-1369-5/13/$31.00 ©2013 IEEE
3. 3D HOLOSCOPIC CONTENT PROCESSING

Viewpoint extraction has been used extensively in the literature to develop algorithms for processing of 3D holoscopic images [1]. The basic idea of the viewpoint extraction is shown in Figure 2. A single viewpoint image is constructed by extracting one pixel from each lens. For example, assembling of every first pixel (1, 1) of each lens generates the viewpoint image (1, 1). Viewpoint image (n, n) is formed by selecting the pixel (n, n) under each lens to generate. Nine different viewpoints are selected respectively as shown in Figure 3.

![Figure 2. Viewpoint extraction from a 3D holoscopic image, (M, N): the number of MLA, (n, n): the pixels under a single microlens. The total of n² viewpoints can be generated. The resolution is M×N for each viewpoint image.](image)

4. 3D HOLOSCOPIC VIDEO CODING USING SELF-SIMILARITY COMPENSATED PREDICTION

Due to the large amount of data associated to holoscopic pictures, efficient coding algorithms are needed to transmit this type of content over actual networks. It is well-known the superior performance of the High Efficiency Video Coding (HEVC) standard for high-quality video coding compared with prior standards. Nevertheless, further performance improvements can be obtained for 3D holoscopic content by exploring the inherent redundancy of this type of content. In this sense, the coding scheme described in this paper attempts to exploit the significant self-similarity (SS) existing between neighboring micro-images, due to the small angular disparity between adjacent microlenses, for improving coding efficiency.

4.1 Self-Similarity Compensated Prediction

A scheme for self-similarity estimation and compensation was first proposed in [14] in order to explore the high self-similarity between neighboring micro-images in a given holoscopic picture and improve the performance of the H.264/AVC. Similarly to motion estimation, the self-similarity estimation process uses block-based matching in order to find the best prediction — in terms of a suitable matching criterion — for a given picture block to be encoded. However, in this case, the search area is restricted to the already coded and reconstructed area of the picture being encoded. This previously coded area forms the SS reference, which is continuously updated as the picture is encoded. As a result, the chosen block becomes the candidate predictor and the displacement between the two blocks is encoded as a vector (similarly to a motion vector). This vector is referred to as the Self-Similarity Vector (SSV). This prediction process has the advantage of being able to find efficient predictions for 3D holoscopic content independently of the structure of the underlying micro-lens array, and consequently, the arrangement of the micro-images.

In the self-similarity compensation block, the inverse quantized and inverted transformed prediction residual is added to the predictor to form the reconstructed data that is stored in the prediction memory to be available for future predictions.

4.2 Self-Similarity Prediction in Intra-coded Frames

HEVC [15], uses as the basic unit for compression a square block of 2N×2N samples with a size up to 64×64, called a Coding Unit (CU). CUs can be recursively split into four smaller CUs until a pre-defined minimum size is reached (i.e., 8×8). Each CU contains one or multiple rectangular Prediction Units (PU), which correspond to the basic units that carry information related to the prediction modes. In a self-similarity compensation sense, better predictions can be obtained as more flexible the block partition scheme can be used (in terms of block sizes and shape patterns). In this sense, two possible prediction modes are defined and referred to as SS and SS-skip prediction modes. The SS prediction mode – based on the HEVC Inter prediction mode [15] – uses the eight HEVC PU partition patterns (i.e., 2N×2N, N×2N, 2N×N, N×N, 2N+nU, 2N+nD, nL×2N and nR×2N), and the SS estimation and compensation process instead of the motion estimation and compensation.

For each PU a self-similarity vector prediction is computed, in a Rate-Distortion Optimization (RDO) sense, using a scheme derived from the HEVC Advanced Motion Vector Prediction (AVMP) [15]. The SS-skip prediction mode is derived from the HEVC merge technique [15], used in the HEVC Inter prediction, where the motion vector is inferred from multiple candidates formed by motion vectors of 5 spatially neighboring PUs, as well as a temporally neighboring PUs. However, the modified merge scheme used in the SS-skip mode limits the SS vector candidates from only spatially neighboring PUs, and the referencing is limited to the area defined in the self-similarity estimation process. This guarantees that the CU signaled by the chosen self-similarity vector is already available as a self-similarity reference at decoding time.

4.3 Self-Similarity Prediction in Inter-coded Frames

In HEVC, each reference picture used for prediction is distinguished by the Picture Order Count (POC) difference in the reference picture list. In order to enable the SS prediction modes in...
Inter coded frame, a POC difference of zero is allowed in the semantics of the slice header so as to access the SS reference picture. Because there is only one SS reference picture for each picture itself, no additional index is necessary to distinguish between a SS reference other temporal reference pictures.

Consequently, no significant changes are needed for the reference picture list construction. This list is initialized as for a conventional 2D video and may include any available temporal reference pictures besides the SS reference, which is appended to the list becoming available for prediction of the current picture.

In terms of PU level, the decoding modules of the proposed coding scheme do not need to be aware of whether a reference picture is temporal or SS reference picture, this distinction is done in the reference picture list parameters. Consequently, the prediction of a 3D holoscopic picture becomes adaptive, in the sense that the encoder can select the best reference among temporal and SS reference pictures in a RDO sense, resulting in only an index for the reference picture list.

### 4.4 3D Holoscopic Video Coding Performance Evaluation

This section assesses and analyses the rate-distortion (RD) performance of the proposed self-similarity compensated prediction scheme for 3D holoscopic video coding. For this, the proposed codec based on the HEVC reference software HM-9.1 enhanced with the proposed prediction modes – referred to as HEVC+SS – is compared against the HEVC reference software version HM-9.1 – referred to as HEVC (HM9.1).

Two 3D holoscopic test sequence called Demichelis Cut and Demichelis Spark with full parallax, spatial resolution of 2880×1620 and 150 frames at 25 Hz have been used to assess the performance of the proposed coding scheme (see Figure 4).

The test sequences have been encoded according to the common test conditions defined in JCTVC-K1100 [16] for “Low-delay, main, P slices only” configuration, considering an “IPPP…” structure with: i) Group of Pictures (GOP) of fixed size equal to 1 and ii) Intra refresh period equal to 9. Four rate-distortion points corresponding to quantization parameter (QP) values (22, 27, 32 and 37) have been considered and a full search with search range of 128 was adopted. The rate-distortion performance is measured in terms of the average luminance Peak Signal-to-Noise Ratio (PSNR Y) and the Bjontegaard Delta (BD) measurement method [9] in terms of PSNR and bitrate (BR).

An additional analysis was performed due to the observation that using the SS prediction scheme in Intra-coded slices results in a significantly lower amount of coded bits with a slight decrease in the average PSNR in comparison with HEVC for the same QP. As can be seen in Figure 5, this PSNR decrease in the first Intra-coded frame of the HEVC+SS affects significantly the consecutive Inter-coded frames, which also present lower PSNR when compared to the HEVC (HM9.1). This shows that, it is convenient to properly adjust the QP value for these modified Intra-coded frames. Therefore, two QP adjustments (only for the modified Intra-coded frames) are also considered in the RD performance results:

i) **HEVC+SS (QP-adjust. by Distortion)** — in this case an offset for the QP values used in the modified Intra-coded frames was estimated so as to result in a PSNR value similar to the PSNR achieved for HEVC Intra-coded frames;

ii) **HEVC+SS (QP-adjust. by Rate)** — in this case an offset for the QP values used in the modified Intra-coded frames was estimated so as to result in a bitrate value similar to the bitrate achieved for HEVC Intra-coded frames;

In the abovementioned QP-adjustment cases, the range of offset values varies from -1 to -5.
coding scheme in the aforementioned configuration. It can be seen that introducing the SS prediction modes into Inter-coded frames is advantageous to improve the global coding efficiency of 3D holoscopic video content. It is important to notice, however, that this is particularly evident for the lower bit rates, where the gains go up to 1.2 dB as shown in Figure 6 and Figure 7.

<table>
<thead>
<tr>
<th>Bjontegaard Delta (BD)</th>
<th>Demichelis Cut</th>
<th>Demichelis Spark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR [dB]</td>
<td>BR [%]</td>
</tr>
<tr>
<td>HEVC+SS (QP-adjust by Distortion)</td>
<td>0.28, -14.74</td>
<td>0.35, -18.71</td>
</tr>
<tr>
<td>HEVC+SS (QP-adjust by Rate)</td>
<td>0.28, -15.11</td>
<td>0.36, -18.26</td>
</tr>
<tr>
<td>HEVC+SS (QP-adjust by Rate)</td>
<td>0.25, -8.39</td>
<td>0.30, -8.96</td>
</tr>
</tbody>
</table>

Figure 6 RD performance for Demichelis Cut test sequence

Figure 7 RD performance for Demichelis Spark test sequence

5. CONCLUSIONS

This paper presented some novel achievements in 3D holoscopic content acquisition, processing, and coding. With respect to 3D holoscopic content acquisition and processing, a flexible single aperture video capturing setup has been presented and the main issues in terms of content acquisition and processing were analyzed. In terms of 3D holoscopic video coding, the performance of the proposed self-similarity compensated prediction technique was shown to outperform HEVC, notably for lower bit rates. Moreover, it was also shown that the gains obtained for Inter coding can be further improved if an adequate QP adjustment for Intra-coded frames is done.

6. ACKNOWLEDGEMENTS

This work was supported by the European Commission under the FP7 project 3D VIVANT (Live Immersive Video-Audio Interactive Multimedia), and by the FCT (Fundaçao para a Ciência e a Tecnologia, Portugal). The authors acknowledge the 3D VIVANT project for providing the 3D holoscopic sequence Demichelis Cut and Demichelis Spark.

7. REFERENCES