SPATIAL PREDICTION BASED ON SELF-SIMILARITY COMPENSATION FOR 3D HOLOSCOPIC IMAGE AND VIDEO CODING

Caroline Conti[1], *João Lino*[1,2], *Paulo Nunes*[1,3], *Luís Ducla Soares*[1,3], *Paulo Lobato Correia*[1,2]

[1]Instituto de Telecomunicações, [2]Instituto Superior Técnico, [3]ISCTE - Instituto Universitário de Lisboa

E-mail: {caroline.conti, joao.lino, paulo.nunes, luis.soares, paulo.correia}@lx.it.pt

ABSTRACT

Holoscopic imaging, also known as integral imaging, provides a solution for glassless 3D, and is promising to change the market for 3D television. To start, this paper briefly describes the general concepts of holoscopic imaging, focusing mainly on the spatial correlations inherent to this new type of content, which appear due to the micro-lens array that is used for both acquisition and display. The micro-images that are formed behind each micro-lens, from which only one pixel is viewed from a given observation point, have a high cross-correlation between them, which can be exploited for coding. A novel scheme for spatial prediction, exploring the particular arrangement of holoscopic images, is proposed. The proposed scheme can be used for both still image coding and intra-coding of video. Experimental results based on an H.264/AVC video codec modified to handle 3D holoscopic images and video are presented, showing the superior performance of this approach.

Index Terms — Integral imaging; holoscopic images and video; micro-lenses; 3D video; spatial prediction.

1. INTRODUCTION

Users continuously demand richer, more immersive and closer to reality viewing experiences. After the introduction of color displays and high definition images, 3D video promises to be the next revolution in visual technology. The recent momentum in the production of 3D content for cinema applications is a good example that the revolution has started.

However, creating a truly realistic 3D viewing experience in an ergonomic and cost effective manner is a fundamental engineering challenge, which still has not been solved. The methods for creating and displaying 3D images that are already established, such as those based on light polarization or active-shutter techniques, have several drawbacks. These include: i) requiring the viewer to wear special glasses in order to create the depth perception [1]; ii) not providing motion parallax (i.e., when the viewer moves the viewpoint remains the same); and iii) causing eye strain, headaches and in some cases nausea. Holoscopic imaging, also known as integral imaging, however, allows 3D images to be viewed without any special eyewear, exhibiting continuous motion parallax throughout the viewing zone, presenting a variety of many different views, depending on the observers' position [2][3][4], with no discomfort. In addition to these advantages over current 3D systems, holoscopic systems are also suited for multi viewer applications and can be obtained with a single aperture camera and then be reproduced on a conventional flat panel display, both with an overlaid micro-lens array. In fact, due to these advantages and the fact that it is now becoming practical, 3D holoscopic imaging technology is now accepted as a strong candidate for next generation 3D television [5].

In order to transmit 3D holoscopic images and video on a limited-bandwidth network with an acceptable quality, efficient coding tools are needed, that fully exploit the new inherent spatial and temporal correlations existing in this type of data.

Several coding schemes have already been proposed in the literature for 3D holoscopic still images, which can also be used for intra-coding video frames. Most of these coding schemes are based on the three-dimensional discrete cosine transform (3D-DCT) [3][6][7][8]. These schemes take advantage of the existing redundancy within the micro-images (i.e., images formed behind each micro-lens), as well as the redundancy between adjacent micro-images, by applying the 3D DCT to a stack of several micro-images. This is possible because 3D holoscopic images are inherently divided into small non-overlapping blocks referred to as micro-images.

Other schemes, in alternative to the DCT, rely on the discrete wavelet transform (DWT) [9][10]. For instance, in [9], the various sub-images are extracted from the 3D holoscopic image, by extracting one pixel with the same position from each micro-lens. Each of these sub-images will then be decomposed using a 2D DWT, resulting in an array of coefficients corresponding to several frequency bands. The lower frequency bands of the sub-images are assembled and compressed using a 3D DCT followed by Huffman coding, while the remaining higher bands are simply quantized and arithmetic encoded.

For video, not many schemes have been proposed in the literature, examples being those in [11][12]. In these papers, the authors propose to decompose the 3D holoscopic video sequence into several sub-image video sequences and then jointly exploit motion (temporal prediction) and disparity between adjacent sub-images to perform compression. In these schemes, the spatial redundancy is obviously exploited by the disparity estimation part of the scheme, similarly to what is done in multiview video coding (MVC) [13]; as such, a precise knowledge of the image structure is needed, notably the sub-image dimensions.

This paper proposes a novel scheme for spatial prediction, exploring the particular arrangement of holoscopic images, notably the intrinsic self-similarity of this type of images, without requiring explicit knowledge of how the micro-images are arranged and their size. The proposed scheme can be used for both still image coding and intra-coding of video.

The remainder of the paper is organized as follows. Section 2 briefly explains the general concepts and structure of holoscopic content, in order for the reader to better understand how the spatial prediction will be made. Section 3 describes the proposed spatial prediction scheme and Section 4 performs the evaluation of that scheme. Finally, Section 5 concludes the paper.

2. SPATIAL RELATIONSHIPS IN 3D HOLOSCOPIC IMAGES AND VIDEO FRAMES

3D holoscopic imaging, first proposed by G. M. Lippmann in 1908 [4], relies on the use of a series of micro-lenses at the picture surface to create the impression of depth and provide the user with full motion parallax without requiring special glasses. The process uses an array of small spherical micro-lenses, known as a "fly's eye" lens array, to both record and display the 3D holoscopic image (see Figure 1). At the display side, this array is designed so that when viewed from different angles, different images will be visible. This happens because, for a given direction, only one pixel from each micro-image is visible by the user. This technology can be used for creating 3D images on a flat panel display. If motion pictures are considered, instead of still images, this results in 3D holoscopic video. At the capture side, the operation is basically the same, but the micro-lens array is applied to the sensor.

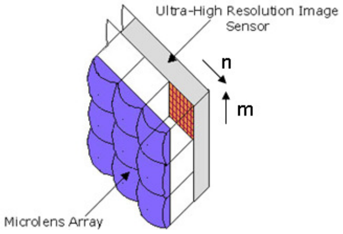

Figure 1 – Example of a micro-lens array applied to an image sensor for capturing holoscopic content

Behind the micro-lens array, the planar intensity distribution representing a holoscopic image consists of a 2D array of micro-images of m×n pixels, due to the structure of the micro-lens array used in the capture, and therefore it could be simply encoded by any 2D image or video encoder. However, each micro-lens works as an individual small low-resolution camera, recording a different image of the same scene, at slightly different angles. Due to the small angular disparity between adjacent micro-lenses, a significant cross-correlation exists between neighboring micro-images and, therefore, this can be exploited for improving coding efficiency. Additionally, a significant correlation also exists between neighboring pixels within each micro-image. These two types of correlation are clearly visible in Figure 2.

Unidirectional holoscopic imaging is a special case of what has just been described, where a 1D cylindrical lens array is used for both capture and display. The resulting 3D image only exhibits horizontal parallax, with the captured image consisting of one pixel wide vertical lines, one line for each viewing position.

The proposed spatial prediction approach, described in the next section, is especially suited for full-parallax holoscopic content, but can also be applied when only unidirectional parallax is considered.

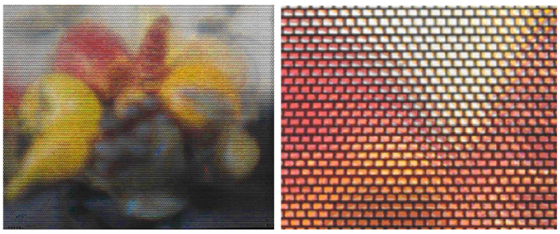

Figure 2 – Example of a 3D holoscopic image: (left) Fruit test image; (right) Test image enlargement, showing the inherent holoscopic spatial structure due to the micro-lens array

3. PROPOSED SPATIAL PREDICTION SCHEME BASED ON SELF-SIMILARITY COMPENSATION

The type of spatial correlation present in 3D holoscopic content can be seen as a kind of self-similarity. As such, it is something that an encoder should exploit in order to improve the coding efficiency.

The use of the term self-similarity in the context of this paper should not be confused with the self-similarity in the context of fractals (and fractal-based image coding), where it basically refers to scale-invariance [14].

3.1. Proposed Codec Architecture

The proposed spatial prediction scheme corresponds to a module that fits in the architecture of a 3D holoscopic image/video codec, as depicted in Figure 3, which is based on the H.264/AVC architecture [15]. This scheme introduces a new spatial prediction mode, in addition to all the existing H.264/AVC Intra modes. The new Intra mode (INTRA_SS) uses a self-similarity estimation and a self-similarity compensation block, to find the best predictor for the current block being encoded in the area of the current image/frame that has already been encoded (and reconstructed, since this is what will be available at the decoder).

3.2. New Self-Similarity Spatial Prediction (SSSP) Mode

For each 16×16 macroblock to be encoded, the self-similarity estimation module uses block-based matching, in an area of the previously coded and reconstructed macroblocks of the current picture, to find a full-pel region with the best match for prediction of the current macroblock. The chosen 16×16 block becomes the candidate predictor for the current macroblock being encoded. When the self-similarity prediction mode becomes the best mode, in a rate-distortion sense, from the set of all possible macroblock intra-coding modes, the relative position between the two macroblocks is encoded and transmitted as a vector (similarly to a motion vector). This vector is here referred to as the self-similarity vector. In addition to this, the prediction residual is also encoded and transmitted. The allowed search area for self-similarity estimation is illustrated in Figure 4.

In the self-similarity compensation block, the inverse quantized and inversed transformed prediction residual is added to the predictor to form the reconstructed macroblock that is stored in the prediction memory in order to be used for prediction of future macroblocks.

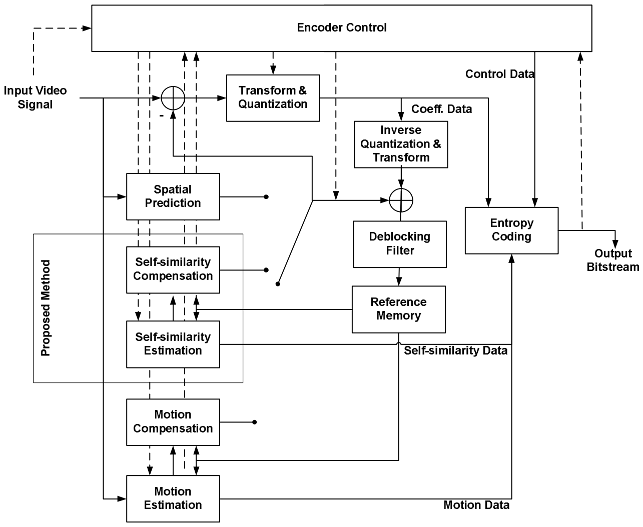

Figure 3 – Proposed 3D holoscopic codec architecture, including the self-similarity spatial prediction scheme

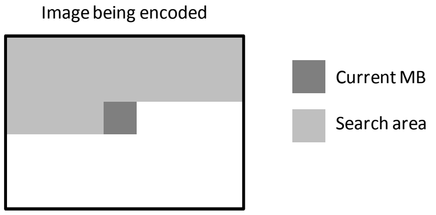

Figure 4 – Illustration of the allowed search area for the estimation of the self-similarity vector

3.3. Mode Decision

The decision to choose the best macroblock mode is accomplished through the rate-distortion optimization (RDO) technique, where the best macroblock mode is selected by minimizing the following Lagrangian cost function:

$$J_{MODE} = D(MODE,QP) + \lambda_{MODE} \times R(MODE,QP) \qquad (1)$$

where MODE is one of the allowed Intra macroblock coding modes (i.e., INTRA 4×4, INTRA 16×16, INTRA 8×8 or the new INTRA SS), QP is the macroblock quantization parameter, and $D(MODE,QP)$ and $R(MODE,QP)$ are, respectively, the distortion (between the original and the reconstructed macroblock) and the number of bits that will be achieved by applying the corresponding MODE and QP. λ_{MODE} is the Lagrange multiplier parameter and is computed as in [13].

Therefore, to encode a given macroblock, the proposed scheme computes the best Intra mode RD cost, J_{INTRA}, from the set of all possible Intra modes by using Equation (1), where

$$J_{INTRA} = \min_{MODE \in \mathbf{S}_{INTRA}} \left(J_{MODE} \right) \qquad (2)$$

where $\mathbf{S}_{INTRA}$ is the set of allowed Intra modes. The best Intra mode is the one with the lowest Intra RD costs.

4. PERFORMANCE EVALUATION

This section evaluates the performance of a modified H.264/AVC codec with the additional proposed SSSP mode against the normative H.264/AVC codec using typical settings for intra-coding. In these tests, the deblocking filter has been disabled.

For these tests, the *Fruit, Plane* and *Toys* holoscopic test images, with full parallax, have been used (see Figure 5).

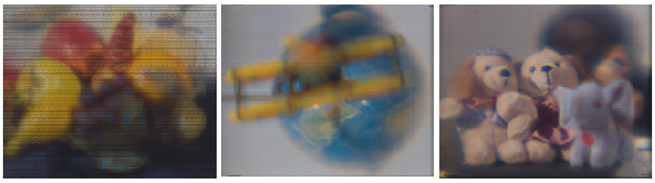
(a) (b) (c)

Figure 5 – Test images: (a) Fruit; (b) Plane; (c) Toys

For the SSSP mode decision, only 16×16 blocks are currently considered. In terms of search pattern, an exhaustive block-based estimation with integer pixel positions was used. This way, in the modified codec, a total of four modes were available to encode each macroblock: Intra 16×16, Intra 8×8, Intra 4×4 and the Intra SS mode.

Table 1 shows the percentage of each Intra coding mode for each test image encoded with a QP of 31. The macroblock mode decisions were done as described in Section 3.3. As can be seen in Table 1, the most frequent mode for the *Fruits* image is clearly the new proposed Intra SS mode (i.e., 53%). For the other two images, the percentages are not as high, but are still quite significant. This is clearly reflected in the rate-distortion curves of Figures 6-8.

Table 1 – Relative Intra mode selection statistics

Test Image (Resolution)	Intra SS	Intra 16x16	Intra 8x8	Intra 4x4
Fruit (680×562)	53 %	10 %	36 %	1 %
Plane (698×570)	36 %	21 %	38 %	4 %
Toys (698×570)	27 %	28 %	39 %	6 %

As can be seen in Figure 6-8, the modified codec including the new Intra SS mode always outperforms the standard H.264/AVC codec, in terms of the achieved PSNR Y values. The achieved gains can go up to 3.4 dB for the *Fruit* image, 1.0 dB for the *Plane* image and 0.5 dB for the *Toys* image. These gains are closely related to the relative percentage of Intra SS mode usage.

5. FINAL REMARKS

This paper proposes a novel spatial prediction scheme that exploits the self-similarity inherent to 3D holoscopic content. The proposed method can be used for both still image coding, as well as intra-coding of video. With this scheme, a better spatial prediction was obtained for full-parallax holoscopic content, leading to an improved coding efficiency with respect to H.264/AVC of up to 3.4 dB. The proposed SSSP scheme has also been tested with unidirectional holoscopic content, but the improvements were not significant; improving the performance for this content is left as future work.

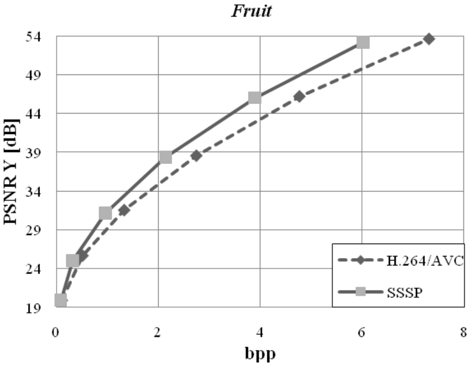

Figure 6 – PSNR results for the Fruit image

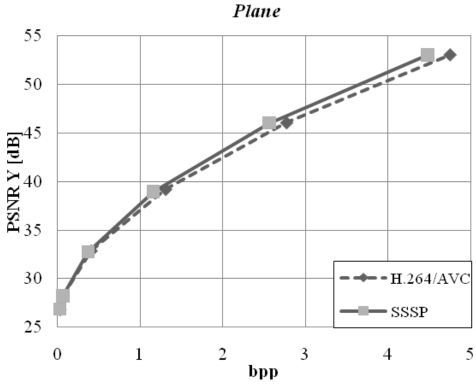

Figure 7 – PSNR results for the Plane image

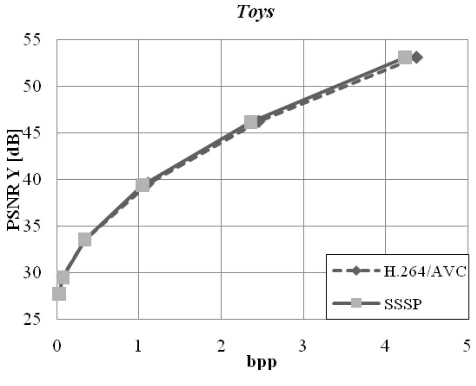

Figure 8 – PSNR results for the Fruit image

ACKNOWLEDGEMENTS

The authors acknowledge the support of the European Commission under the FP7 project 3D VIVANT (Live Immerse Video-Audio Interactive Multimedia).

REFERENCES

[1] Y. Zhu, T. Zhen, "3D Multi-View Autostereoscopic Display and Its Key Technologie,", *Proc. of the Asia-Pacific Conference on Image Processing (APCIP 2009)*, vol. 2, pp. 31-35, Shenzhen, China, July 2009.

[2] G. Milnthorpe, M. McCormick, N. Davies, "Computer Modeling of Lens Arrays for Integral Image Rendering", *Proc. of Eurographics UK Conference*, Leicester, UK, June 2002.

[3] R. Zaharia, A. Aggoun, M. McCormick, "Adaptive 3D-DCT Compression Algorithm for Continuous Parallax 3D Integral Imaging", *Signal Processing: Image Communication*, vol. 17, no. 3, pp. 231-242, March 2002.

[4] G. Lippmann, "Epreuves Reversibles Donnant la Sensation du Relief", *Journal de Physique Théorique et Appliquée*, vol. 7, no. 1, pp. 821-825, November 1908.

[5] M. Okui, F. Okano, "3D Display Research at NHK", *Workshop on 3D Media, Applications and Devices*, Berlin, Germany, October 2009.

[6] A. Aggoun, "A 3D DCT Compression Algorithm For Omnidirectional Integral Images", *Proc. of the IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP 2006)*, vol. 2, pp. 517-520, Toulouse, France, May 2006.

[7] R. Zaharia, A. Aggoun, M. McCormick, "Compression of Full Parallax Colour Integral 3D TV Image Data Based on Subsampling of Chrominance Components", *Proc. of the IEEE Data Compression Conference (DCC 2001)*, pp. 27-29, Snowbird, UT, USA, March 2001.

[8] M. C. Forman, A. Aggoun, "Quantisation Strategies for 3D-DCT based Compression of Full Parallax 3D Images", *Proc. of the IEE International Conference on Image Processing Applications (IPA 1997)*, Dublin, Ireland, pp. 32-35, July 1997.

[9] A. Aggoun, M. Mazri, "Wavelet-based Compression Algorithm for Still Omnidirectional 3D Integral Images", *Signal, Image and Video Processing*, vol. 2, no. 2, pp. 141-153, June 2008.

[10] E. Elharar, A. Stern, O. Hadar, B. Javidi, "A Hybrid Compression Method for Integral Images Using Discrete Wavelet Transform and Discrete Cosine Transform", *Journal of Display Technology*, vol. 3, no. 3, pp. 321-325, September 2007.

[11] S. Adedoyin, W. A. C. Fernando, A. Aggoun, "A Joint Motion and Disparity Motion Estimation Technique for 3D Integral Video Compression Using Evolutionary Strategy", *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 732-739, May 2007.

[12] S. Adedoyin, W. A. C. Fernando, A. Aggoun, "Motion and Disparity Estimation with Self Adapted Evolutionary Strategy in 3D Video Coding", *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1768-1775, November 2007.

[13] Joint Video Team, "Joint Multiview Video Coding 8.3.1", November 2010.

[14] B. Mandelbrot, "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension", *Science, New Series*, vol. 156, no. 3775, pp. 636-638, May 1967.

[15] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, July 2003.