Manuscript for Review

Compressing Depth Maps using Multiscale Recurrent Pattern Image Coding

Journal:	*Electronics Letters*
Manuscript ID:	ELL-2010-0135
Manuscript Type:	Letter
Date Submitted by the Author:	14-Jan-2010
Complete List of Authors:	Graziosi, Danillo; COPPE/UFRJ, Programa de Engenharia Elétrica Rodrigues, Nuno; Instituto Politécnico Leiria
Keywords:	IMAGE COMPRESSION, DIGITAL SIGNAL PROCESSING, IMAGE CODING, IMAGE PROCESSING, PATTERN MATCHING, SIGNAL PROCESSING, STEREO IMAGE PROCESSING

Compressing Depth Maps using Multiscale Recurrent Pattern Image Coding

Danillo B. Graziosi, Nuno M. M. Rodrigues, Carla L. Pagliari, Sergio M. M. de Faria, Eduardo A. B. da Silva and Murilo B. De Carvalho

The use of the Multidimensional Multiscale Parser (MMP) algorithm for depth maps coding is proposed. The compression method uses a block-based approach, where efficient prediction combined with pattern matching is applied to the encoding of grayscale images, which convey the disparity or depth information of a 3D image. Simulations' results show gains of up to 10dB when compared with state-of-the-art methods, such as JPEG2000 and H.264/AVC.

Introduction: With the development of new equipments with tri-dimensional (3D) capabilities and the growing availability of 3D and multiview media, the interest in 3D image coding has grown in the past few years. However, this exciting new technology comes at a cost. The additional views impose a higher bit rate, thus several proposals for efficient 3D coding have been made. A popular approach is to encode only a reduced set of views along with their disparity/depth information. The missing view(s) can then be reconstructed at the receiver side using the decoded base view(s) and the depth map.

The depth information can be quantized and represented as a grayscale image. Depth values within the same object tend to be similar, while those belonging to different objects are usually quite different and may abruptly change at objects' boundaries. The preservation of sharp discontinuities present in the disparity/depth map is very important in the reconstruction of the missing view, as well as in assuring precise rendering at the 3D monitor. Therefore, the coding of this information represents an important challenge for 3D image coding.

In general, the encoding of the depth data is performed by using a standard compression algorithm and some technique to preserve the sharp edges. Sharp edges have information associated with high frequencies, which are usually eliminated by these encoders. For example, in [1] a region of interest is used to prevent JPEG2000 compression artifacts, and in [2] filtering techniques improve the performance of an H.264/AVC encoder.

From the above, one can infer that a good encoder for depth information should preserve sharp edges, while efficiently coding smooth areas. In this Letter we show that the Multidimensional Multiscale Parser algorithm (MMP) [3,4] is well suited for the encoding of depth data. This is so because it makes no assumption about the input signal. MMP preserves the high frequency components using approximate pattern matching techniques and dictionary updating with patterns from the image itself.

Our simulation results show that MMP outperforms JPEG2000 and H.264/AVC INTRA for encoding depth maps at all bitrates. Furthermore, it preserves object boundaries, which are very important features of these maps.

MMP algorithm: Firstly, the image is segmented into blocks, and then for each block a prediction step is applied, using prediction modes similar to the ones of the H.264/AVC standard [3]. The residue is then encoded by using approximate pattern matching with codewords from a dictionary having elements at different scales. In order to achieve this, an appropriate scale transformation is performed on the dictionary elements prior to the block matching. A code-vector index is then used to represent the original residue. If the match of the residue is not satisfactory, the encoder can segment the block either in the horizontal or the vertical direction (a flag is sent indicating the segmentation direction), and the procedure of approximate pattern matching is replicated for each half of the block (in fact, a new prediction can be performed for each half of the block as well). All the elements (flags, prediction modes and dictionary indexes) are encoded using an adaptive arithmetic encoder.

After coding the block, the reconstructed pattern is inserted in the dictionary. Therefore, the dictionary grows with patterns from the residue image itself, in a Lempel-Ziv fashion. The choices of the segmentation and prediction modes, as well as the codewords that represent the residues, are chosen based on a Lagrangian cost function. Details of this encoding procedure as well as a survey on techniques to enhance dictionary adaptation can be found in [3].

MMP makes no assumption about the source. Instead, it progressively builds its dictionary with codewords drawn from the reconstructed image. Therefore, MMP has a universal character and is well suited for both types of images, smooth and those with strong high frequency content. This contrasts with transform-based encoders, that perform well for smooth images but tend to heavily quantize high-frequency patterns. In MMP, once a high-frequency pattern occurs, it is included in the dictionary, enabling MMP to efficiently encode similar subsequent blocks. In addition, MMP's prediction step is able to effectively exploit the correlation in smooth image areas. Latest results for the MMP compression performance show that it consistently outperforms state-of-the-art encoders such as JPEG2000 and H.264/AVC for smooth and non-smooth images (e.g. text and graphics). Further details can be found in [4].

Experimental results: Here we present simulation results for two images (Tsukuba and Cones) taken from the Middlebury test set[1] and the first frame of the sequence "Book Arrival". The disparity map of the first two images corresponds to the ground-truth, and objects with different depth are represented by sharp edges. The depth map of the first frame of views 2 and 4 for the "Book Arrival" sequence presents blurred edges and corresponds to a much smoother image. It is important to stress that this image was obtained using the reconstruction technique described in [5], hence is not a ground truth. The disparity/depth information was compressed with JPEG2000, H.264/AVC INTRA and MMP. Software JM 16.2 was used (high profile, level

[1] http://vision.middlebury.edu/stereo/

4.0) for H.264 coding. The images were intra encoded with a QP varying from 10 to 50. For the JPEG2000, the *Kakadu* software was used, with Qstep parameter varying from 0.7 to 0.005. The MMP software is available online at http://www.lps.ufrj.br/profs/eduardo/mmp/.

For an objective evaluation, Figs. 1 to 3 show the rate-distortion curves of the encoded disparity maps. In the case of disparity maps that present strong edges (Tsukuba and Cones), MMP outperforms both JPEG2000 and H.264/AVC in all rates, as shown in Fig. 1 and 2. For smoother images, like those of the “Book Arrival” sequence, we can notice (Fig. 3) that the MMP's rate-distortion performance is equivalent to H.264/AVC and slightly better than JPEG2000, the state-of-the-art codecs for this type of images.

Fig. 4 shows details of the disparity map of stereo pair Tsukuba encoded with JPEG2000, H.264 and MMP. It is possible to notice ringing artifacts produced by JPEG2000 compression and also scattered pixels around the strong edges as well as blurring of the soft edges produced by H.264/AVC compression. One major concern is the preservation of sharp edges of the disparity maps that, in general, identify the boundaries of objects. When an error occurs in these areas, it can impair the reconstructions of the missing view and/or the 3D rendering. Fig. 4d clearly shows how MMP preserves the strong edges of the objects. The prediction step for block with strong edges usually creates a high frequency pattern residue that is less likely to exist in the dictionary. In this case, MMP recursively segments the original block until a satisfactory match is found. This creates a new pattern, that is then added to the dictionary and can subsequently be used to encode other prediction mismatches more efficiently, allowing the preservation of the strong edges of the objects.

Conclusion: MMP is an efficient alternative for coding disparity/depth maps for 3D systems, when compared with state-of-the-art encoders such as H.264/AVC and JPEG2000.

Acknowledgment: This work has been supported by Fundação Para a Ciência e Tecnologia (FCT), under project PTDC/EEA-TEL/66462/2006/

References

1 ZANUTTIGH, P.; CORTELAZZO, G.M.: "Compression of depth information for 3D rendering", *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2009, 4-6 May 2009 Page(s):1 - 4

2 OH, K.-J.; YEA, S.; VETRO, A.; HO, Y.-S.: 'Depth reconstruction filter for depth coding', *Electronics_Letters,* Volume 45,Issue 6, March 12 2009 Page(s):305 - 306

3 RODRIGUES, N. M. M.; DA SILVA, E. A. B.; CARVALHO, M. B.; DE FARIA, S. M. M.; DA SILVA V. M. M.: "On Dictionary Adaptation For Recurrent Pattern Image Coding". *IEEE Transactions on Image Processing*, 17(9):1640-1653, September 2008

4 GRAZIOSI, D. B.; RODRIGUES, N. M. M.; DA SILVA, E. A. B.; DE FARIA, S. M. M.; DE CARVALHO, M. B.: "Improving Multiscale Recurrent Pattern Image Coding with Least-Squares prediction mode", Proc. IEEE International Conference on Image Processing (ICIP'09), Cairo, Egypt, Nov. 2009.

5. M. TANIMOTO, T. FUJII AND K. SUZUKI, "Multi-view depth map of Rena and Akko & Kayo", ISO/IEC JTC1/SC29/WG11, M14888, October 2007

Authors' affiliations:

Danillo B. Graziosi (Instituto de Telecomunicações, Portugal and PEE/COPPE/DEL/EE, Universidade Federal do Rio de Janeiro, 21941-972, Brazil) (e-mail: danillo@lps.ufrj.br)

Eduardo A. B. da Silva (PEE/COPPE/DEL/EE, Universidade Federal do Rio de Janeiro, 21941-972, Brazil) (e-mail: eduardo@lps.ufrj.br)

Carla L. Pagliari (Department of Electrical Engineering, Instituto Militar de Engenharia, Rio de Janeiro, 22290-270, Brazil) (e-mail: carla@ime.eb.br)
Nuno M.M. Rodrigues and Sergio M. M. De Faria (Instituto de Telecomunicações, Portugal and also ESTG, Instituto Politécnico Leiria, Apartado 4163, 2411-901 Leiria, Portugal) (e-mail: nuno.rodrigues@co.it.pt, sergio.faria@co.it.pt)
Murilo B. De Carvalho (TET/CTC, Universidade Federal Fluminense, Campus da Praia Vermelha, 24210-240, Niterói, Brazil) (email: murilo@telecom.uff.br)

Figure captions:

Fig. 1 Rate-distortion for Cones' disparity map

Fig. 2 Rate-distortion for Tsukuba's disparity map

Fig. 3 Rate-distortion for the disparity map of the first frame of "Book Arrival" sequence

Fig. 4 Encoding details for Tsukuba disparity
a) Original disparity map
b) JPEG2000 (31.70dB @ 0.113bpp)
c) H.264/AVC (40.14dB @ 0.128bpp)
d) MMP (50.05dB @ 0.129bpp)

Figure 1

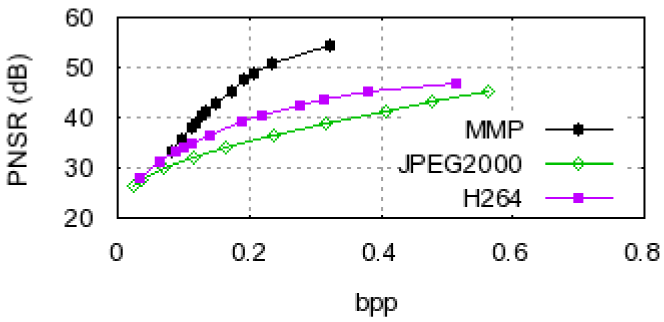


Figure 2

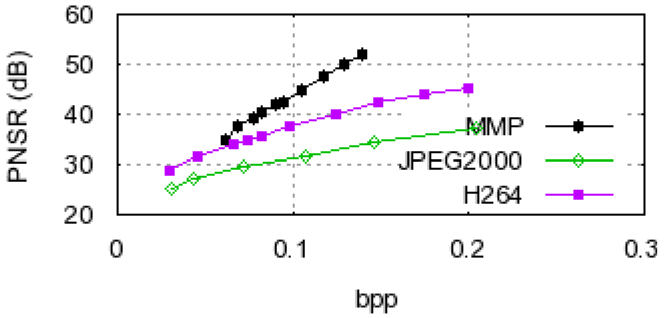


Figure 3

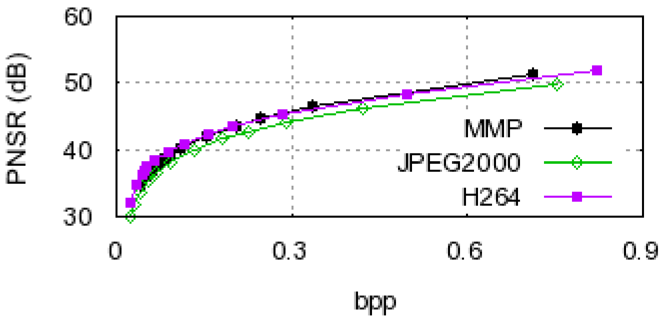
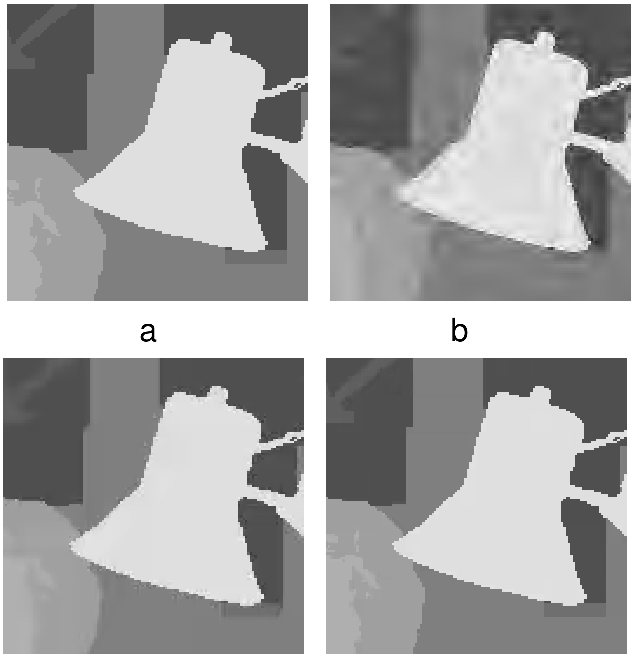


Figure 4



a b

c d