STEREO IMAGE CODING USING MULTISCALE RECURRENT PATTERNS

M. H. V. Duarte [1], *M. B. Carvalho* [2], *E. A. B. da Silva* [3], *C. L. Pagliari* [4], *G. V. Mendonça* [5]

[1,3,5]PEE/COPPE/DEL/EE/Universidade Federal do Rio de Janeiro
Cx.P. 68504, Rio de Janeiro, RJ, Brazil
[2]CTC/TET/Universidade Federal Fluminense
[4]DEE/Instituto Militar de Engenharia
[1]heveline,[3]eduardo,[5]gelson@lps.ufrj.br,[2]murilo@telecom.uff.br,[4]carla@ime.eb.br

ABSTRACT

A stereo image coding method using multiscale recurrent patterns is proposed. The method relies on the matching of recurrent patterns. The input image is segmented in variable-sized blocks. Each segment is coded by making use of contractions, expansions and displacements of elements of a dictionary. The segmentation is ruled by a rate distortion criterion and the dictionary is updated with the concatenation of previously coded elements. The novelty of this work is the absence of the disparity map in the stereo image coding process. That is, the burden of its calculation, the pre- and post-processing phases, the generation of the error images, as well as their coding and transmission are not necessary. In brief, the proposed method cuts off the load of the whole disparity estimation process yet presenting high-quality results at the decoder end.

1. INTRODUCTION

The depth perception offered by stereoscopic video systems can bring realism to many applications such as 3D movies, medicine surgery, video-conferencing, multimedia, and remote operations, among others. However, before the unrestricted use of stereoscopic systems there are some problems to tackle. One challenge is to cope with limited channel bandwidth. This is the main hindrance for making these systems feasible since stereo images require the transmission of, at least, double the amount of data used in monocular image coding systems. The techniques used to code binocular images are similar to those used in monocular image coding. In general, the binocular redundancy between stereo images as well as the temporal redundancy between consecutive frames of each view are exploited in stereo coding schemes [1]. Usually, the reference image (the right-view or left-view) is encoded using a known method like MPEG-2. Next, the disparity map using both views is estimated [2]. The remaining view can be either motion estimated from the previously reference frame, or disparity estimated from the reference view. The resultant error images, the(*Displaced Frame Difference* - DFD (stereo video), and the *Disparity Compensated Difference* - DCD (stereo pairs) are coded and transmitted. The main step to obtain the DCD is the disparity estimation process that tackles an ill-posed problem. The quality of the disparity map rules the amount of information carried by the DCD and the number of bits to code it. Nevertheless, accurate disparity maps demand a high bit rate to be transmitted. One solution is the use of block-based disparity estimators which may produce very inaccurate disparity maps because they are not, in many cases, blockwise constant. Hence, hierarchical block disparity estimation is an improved solution. Larger blocks are used in areas where the disparity varies smoothly, while smaller ones in areas where there is a higher disparity variation. A stereo image pair involves two similar images. Hence, the usage of a recurrent patterns method [3, 4, 5] where the learned patterns in the reference image coding process can be used to code the other image seems to be a good idea. This work employs a new class of coders using multiscale recurrent patterns. A recently proposed method referred to as *Multidimensional Multiscale Parser* - MMP [6, 7] uses contractions and expansions of elements belonging to a dictionary to code each segment of an image. The image is segmented according to a rate distortion criterion and the dictionary is updated with the concatenation of previously coded elements. This method is efficient for monocular image coding, especially when the image is composed by pictures and text. In this paper we propose a coder based on the MMP tailored to stereo image pairs characteristics. In addition to the expansions and contractions of the dictionary elements, displacements of previously coded blocks are used as well. The organization of the paper is as follows. Section 2 starts with a description of the method *Multidimensional Multiscale Parser* - MMP. Next, we describe the modifications implemented to MMP focusing the exploitation of stereo image pairs characteristics. Sections 4 and 5 present, respectively, the experimental results and conclusions.

This work has been supported by CNPq, Brazil, grants n[os] 522381/95-2 and 143217/97-6 and FAPERJ, Brazil, grant n[o] 170.331/2000

2. DESCRIPTION OF THE MMP

In MMP, the image is initially divided in blocks of size $N \times N$. Each block is further segmented in smaller variable-sized blocks. These smaller blocks are then approximated by contracted/expanded versions of matrices in a dictionary. These contractions and expansions are performed using a procedure similar to the usual decimation and interpolation operations [8]. The output of MMP is composed by the dictionary indexes corresponding to each block, as well as information regarding the segmentation. The algorithm starts with a small initial dictionary that is updated as the input data is encoded.

The segmentation in MMP can be represented by a binary tree. An example is depicted in figure 1. The root node of the tree corresponds to an element of dimension $N \times N$. This element can be divided into two smaller elements (children), represented by the two nodes below the root node (Fig. 1). The division, in half, can be horizontal (shown in Fig. 1) or vertical.

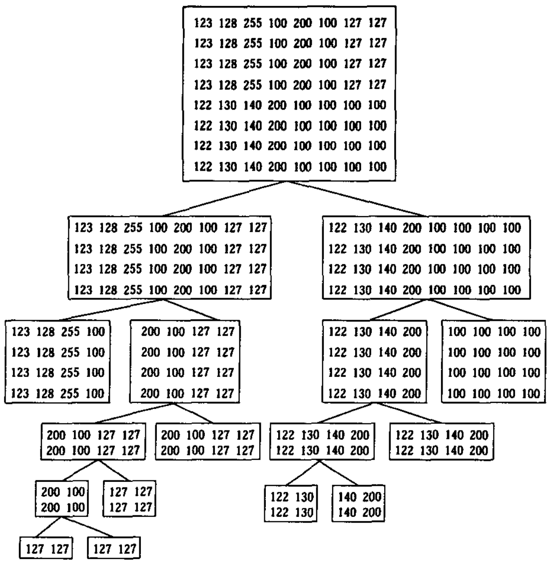

Fig. 1. Binary segmentation tree

Each block corresponding to a leaf (a node without a child) is represented by an index denoting an element of the dictionary. As the elements have variable length, they can be expanded or contracted using a scale transformation, preserving the block dimension. Whenever two leafs of the tree corresponding to two children of the same node are encoded, the resulting element from their concatenation is included in the dictionary. The segmentation tree is obtained by an iterative procedure that attempts to minimize the Lagrangian cost J given by $J = D + \lambda R$ where D is the distortion, R is the rate, and λ is a constant (Lagrange multiplier). The distortion is given by the sum of the mean square error between the image blocks and the corresponding elements of the dictionary. The bit rate is the number of bits per pixel used to transmit the indexes of the elements of the dictionary and the segmentation tree. The segmentation tree is transmitted to the decoder using a sequence of binary flags. The sequence of dictionary indexes and the sequence of binary flags are encoded by an arithmetic coder [9]. The dictionary is updated whenever two leafs connected to the same node are found. These leafs are concatenated and included in the dictionary. As the dictionary grows it is expected that it should contain elements more alike the previously coded blocks, decreasing the number of tree splittings. Consequently, a smaller number of indexes and flags are generated, lowering the bit rate and improving the compression. The decoder starts with the same initial dictionary as the coder. As it receives the flags defining the tree and the elements corresponding to the tree leafs, it starts reconstructing the received image. The dictionary updating process at the decoder is equivalent to that at the encoder. Whenever the decoder receives two terminal nodes they are concatenated and the resultant node is included in the dictionary. The decoding is a much faster operation than the coding process, as the optimization of the segmentation tree is not necessary.

3. STEREO MMP

The purpose of the stereo MMP is to directly encode a stereo image pair, composed by a reference image and a target image, without explicitly evaluating the DCD.

The MMP adaptation is based in two main points:

- The inclusion in the dictionary of elements corresponding to the displacement of previously coded blocks of the reference image;
- The usage of variable-sized blocks, a basic property of MMP. If the cost of using a large block is greater than the cost of using two smaller blocks, the segmentation is performed. Therefore, the displacement can be represented as accurately as needed.

Initially, the reference and the target images are both subdivided in $N \times N$ blocks. Adjacent blocks are grouped to form a row of blocks, or ROB. That is, a ROB is a $N \times M$ slice of the image, where M is its numbers of columns blocks. Each block of the first ROB of the reference image is processed by the standard MMP. After that, the dictionary of MMP has the approximations to all $N \times N$ blocks, as well as the approximations to all sub-blocks that resulted from the segmentation process. Assuming parallel camera conditions, an $n \times m$ block in the first ROB of the target image will correspond to another $n \times m$ block in the first ROB of the reference image, but at a displaced position. The performance of the MMP improves as the matching of blocks becomes more frequent. Therefore, to increase the matching probability, we include displaced versions of the blocks in the dictionary obtained after the processing of the first ROB of the target image. The inclusion of displaced blocks is repeated after the processing of each ROB of the reference image before we encode the corresponding ROB of the target image.

In the stereo MMP, the inclusion of displaced elements in the dictionary replaces the disparity estimation process. When the choice of an element results in a high cost, the decision process that specifies the best segmentation tree decides to halve the block. This is equivalent, in DCD-based methods, to obtain the disparity using smaller blocks, yielding more accurate disparity estimates. Elements corresponding to half-pixel displacements were also included to improve the resolution of the estimation. Parallel-camera conditions are assumed so that disparities are purely horizontal[1], decreasing the number of displaced elements to generate. In addition, this number is limited by a search window given by the minimum and maximum disparity values for the image. The statistics of the usage of the dictionary elements by type show that, when the target image is being coded, the displaced elements are much more used than the others (initial, concatenated, contracted or expanded). Therefore, the dictionary was split in two: one containing the displaced elements, and the other containing all the other types. A flag is used to signal which dictionary was selected. This approach leads to a better estimation of the probability of the dictionary indexes, improving the performance of the arithmetic coder.

4. EXPERIMENTAL RESULTS

The stereo MMP was tested in two versions: the first, that we called StereoMMP_1dic, uses just one dictionary containing all

[1] some images used to test the algorithm were shot with a very small convergence angle, so that the induced vertical disparity can be considered insignificant.

the types of elements (initial, concatenated, expanded, contracted and displaced) to encode both views, and the other, that we called StereoMMP_2dic, uses two dictionaries: one containing initial, concatenated, expanded and contracted elements and another dictionary containing just the displaced elements. The former dictionary is always used to encode the reference image. For encoding the target image (right view) we also use this dictionary, but we have the option to use the latter dictionary (containing only displaced elements), when this offers a smaller cost. A flag signals which dictionary has been chosen. The cost of this flag is also included in the optimization. We can see that, for all the stereo image pairs tested, the stereo MMP using two dictionaries performs better than the one using just one.

The two versions of stereo MMP were run to compress the stereo image pairs CORRIDOR [2], AQUA, SAXO [3] and MAN [4]. Figure 2 shows the rate x PSNR (*Peak Signal to Noise Ratio*) performance results for the pair CORRIDOR. The improvement obtained with the use of two dictionaries is larger for the target image (right view, in figure 2 (b)). This is expected because displaced elements tend to be frequent for the target image. The figure 3 shows the rate x PSNR performance results for the pairs AQUA, SAXO and MAN.

The rate x PSNR performance results of stereo MMP using two dictionaries, including also lower rates, are shown in figure 4 for CORRIDOR. The results of the state-of-the-art [10] are also shown for comparison. We have that, for rates above 0.5 bps, the stereo MMP outperforms the MRF-based hierarchical block matching (HQBM) of [10].

5. CONCLUSIONS

In this paper we proposed the use of multiscale recurrent patterns to encode stereo image pairs. The bit rates of the coded stereo image pairs are reduced by taking advantage of the intrinsic similarity between the two views. Instead of using the widespread approach that uses the disparity map and the disparity compensated differences, this work opens another track to code stereo images. This new path stays away of the high overhead imposed by the disparity estimation phase that aims to produce accurate disparity maps. The algorithm avoids several well-know drawbacks of stereo codecs, presenting efficient alternatives and producing high-quality reconstructed images.

We have developed an efficient coder based on this principle. The analysis of its performance has shown that the inclusion of displaced elements in the dictionary is an effective way of encoding the target images. The algorithm performed very well for image pairs obtained using parallel camera conditions. In addition, its structure allows its adaptation to non-parallel camera geometries. For example, besides the dictionary of displaced blocks used to encode the target frame, one could add dictionaries of elements displaced and distorted taking into account the non-parallel geometry, thus increasing the match probabilities. In brief, the proposed method is quite promising, opening a new avenue in stereo image coding.

[2]Computer Vision and Pattern Recognition Group, University of Bonn

[3]CCETT: Centre Commun d'Etudes de Télédiffusion et Télécommunications (test sequences shot and distributed under RACE DISTIMA European Project), France.

[4]HHI: Heinrich-Hertz-Institut, Berlin, Germany.

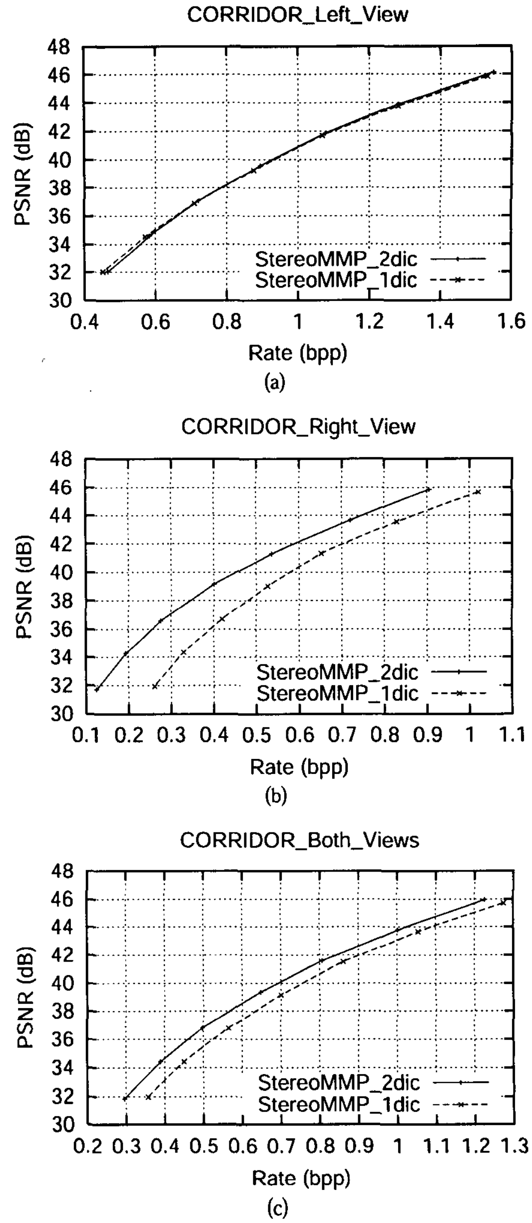

Fig. 2. Rate x PSNR for stereo MMP using just one dictionary including all the elements types (initial, concatenated, contracted, expanded and displaced) to encode both views, and stereo MMP using two dictionaries, one including all the types of elements, except displaced ones, and another including only displaced elements, applied to the stereo image CORRIDOR. (a) left-view, (b) right-view, and (c) both views.

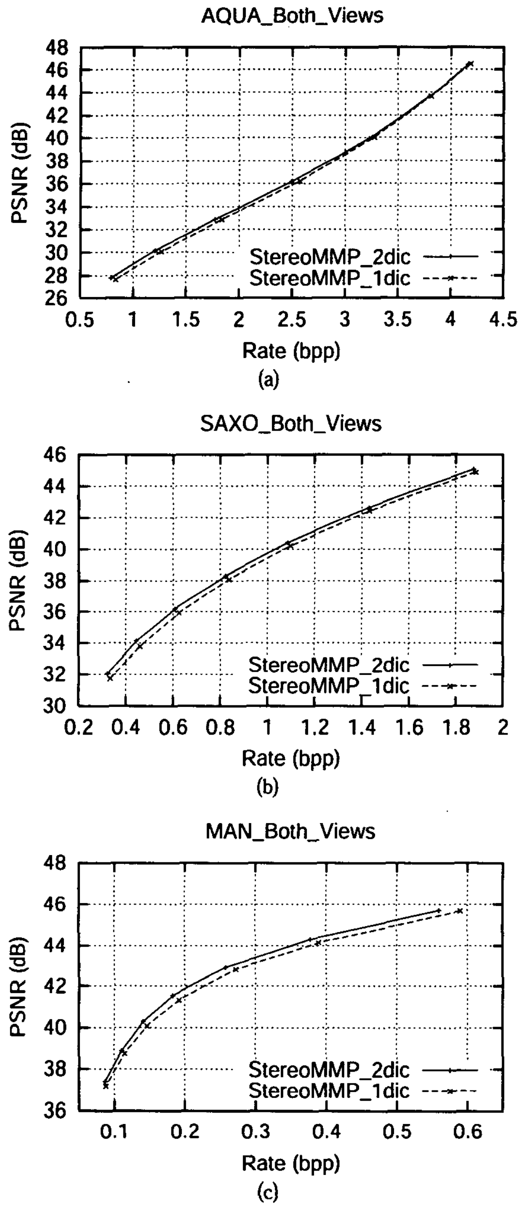

Fig. 3. Rate x PSNR for stereo MMP using just one dictionary including all the elements types (initial, concatenated, contracted, expanded and displaced) to encode both views, and stereo MMP using two dictionaries, one including all the types of elements, except displaced ones, and another including only displaced elements, applied to both views of the stereo images (a) AQUA, (b) SAXO, and (c) MAN.

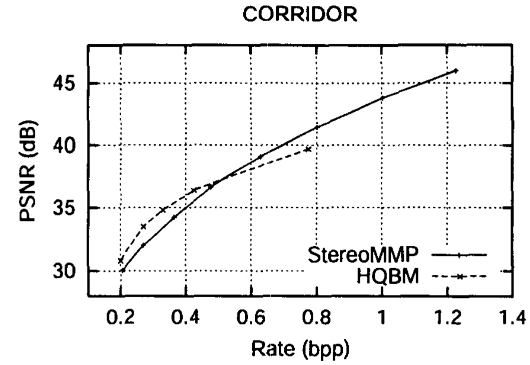

Fig. 4. Rate x PSNR for stereo MMP and HQBM [10] with the stereo image CORRIDOR (ROOM in [10]).

6. REFERENCES

[1] M.G. Strintzis and S. Malasiotis, "Object-based coding of stereoscopic and 3d image sequences: A review," *IEEE Signal Processing Magazine, Special Issue on Stereo and 3D Imag- ing (Invited paper)*, vol. 16, no. 3, pp. 14–28, May 1999.

[2] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, Massachusetts, 1993.

[3] J.Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Transactions on Information Theory*, vol. it-24, no. 5, pp. 530–536, September 1978.

[4] M. Atallah, Y. Genin, and W. Szpankowski, "Pattern matching image compression: algorithmic and empirical results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 618–627, 1999.

[5] M. Alzina, W. Szpankowski, and A. Grama, "2d-pattern matching image and video compression," *IEEE Transactions on Image Processing*, 2001.

[6] M. B. Carvalho, E. A. B. da Silva, and W. A. Finamore, "Rate distortion optimized adaptive multiscale vector quantization," in *2001 IEEE International Conference on Image Processing*, Thessaloniki, October 2001, vol. II, pp. 439–442.

[7] M. B. Carvalho, E. A. B. da Silva, and W. A. Finamore, "Multidimensional signal compression using multiscale recurrent patterns," *Signal Processing - Special Issue on Image and Video Coding Beyond Standards*, 2002, To appear.

[8] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall PTR, Englewood Cliffs, New Jersey, 1995.

[9] T. C. Bell, J. G. Cleary, and I. H. Witten, *Text Compression*, Prentice Hall, Englewood Cliffs, NJ, 1990.

[10] W. Woo, A. Ortega, and Y. Iwadate, "Stereo image coding using hierarquical mrf model and selective overlapped block disparity compensation," in *1999 IEEE International Conference on Image Processing*, Kobe, October 1999.