Stereo Image Coding using Dynamic Template-Matching Prediction

Luís F. R. Lucas*†, Nuno M. M. Rodrigues*†, Eduardo A. B. da Silva‡, Sérgio M. M. de Faria*†
*Instituto de Telecomunicações; †ESTG, Instituto Politécnico de Leiria, Portugal;
‡PEE/COPPE/DEL/Poli, Universidade Federal do Rio de Janeiro, Brazil
e-mails: luisfrlucas@gmail.com, nuno.rodrigues@co.it.pt, eduardo@lps.ufrj.br, sergio.faria@co.it.pt

***Abstract*—Template matching (TM) has been originally proposed as a texture synthesis tool. However, it has been successfully exploited for spatial and inter-frame prediction in video coding. In this paper we investigate the use of TM prediction for stereo image coding.**

In order to efficiently encode stereo images, the TM algorithm was optimized for stereo disparity prediction. Additionally, we have used the state-of-the-art pattern matching based algorithm, the Multidimensional Multiscale Parser (MMP), to encode images predicted with the proposed scheme.

Experimental results of the developed stereo image encoder show that TM is able to efficiently exploit the redundancy between stereo views. The use of the new prediction method with MMP also achieves important coding gains over the state-of-the-art transform based H.264/AVC standard, for stereo image coding.

***Index Terms*—Stereo Image coding, Template Matching, Inter Prediction, Recurrent Pattern Matching**

I. Introduction

Recent advances in multimedia technology employ 3D imaging systems to provide a powerful viewing experience. Due to its greater realism, the depth perception introduced by the stereoscopic systems is being used in many applications. Consequently, the growing use of 3D media content motivates the increasing investigation in stereo image coding.

Stereo vision is generated by presenting a stereo pair to the human observer. The stereo pair refers to two images of the same scene acquired from different viewpoints. The depth is perceived by the human observer when he receives different images in each eye. Stereo pairs use twice the amount of information, so efficient compression schemes that exploit redundancy among views should be used.

Stereo image compression schemes usually encode the reference image, generally assigned to the left view, and exploit the binocular dependency between views to compress the right image. The dependency between views is usually very strong, in spite of the difference between objects' position in the stereo pair, known as disparity. The disparity compensated image can be used as prediction for the right image, and the resulting error image is then coded and transmitted.

Disparity values are usually computed as horizontal shifts between objects in both images of the stereo pair. Vertical component of the disparity is null if the stereo pair is aligned, i.e., the stereo pair was acquired with two cameras with parallel optical axes, or if an image rectification process was applied. Since the stereo disparity is treated as a displacement, most of the disparity computation algorithms may be similar to the motion estimation algorithms employed in inter-frame video coding.

Two main paradigms may be used for motion estimation: explicit or implicit algorithms. In the first group, the estimated displacement between predicted block and target block, given by the motion vector, is explicitly transmitted to the decoder. In the second group of algorithms, motion vectors are not transmitted. Instead, inter-frame motion estimation is implicitly derived based on the causal past.

Popular approaches towards motion or disparity compensation for inter-frame prediction are maily based on explicit algorithms. The most common technique used for stereo image coding is the block-based disparity compensation [1]. This encoding technique acts similarly to the block-based motion compensation, which is applied in video coding. For each block of the image a predicted block is selected by a matching mechanism, using the previously encoded frame. Adaptive block partitioning and fractional displacement compensation are some of the recent improvements introduced in the block-based motion prediction. The state-of-the-art H.264/AVC standard benefits from these approaches to achieve higher gains when encoding both video and stereo data sources.

In this paper a new approach for efficient stereo image coding is presented. We developed a new template matching (TM) predictor suitable for stereo image pairs. The efficiency of the new TM prediction was tested by using it together with a recurrent pattern matching image encoder, known as Multidimensional Multiscale Parser (MMP) [2], [3]. The MMP algorithm is being developed as an alternative coding paradigm to the traditional transform based standards.

Experimental results show that the proposed algorithm is able to consistently outperform the traditional transform-based encoders, using implicit disparity estimation.

This paper is organized as follows. Section II presents the investigation on the proposed TM strategies. Section III analyzes implementation details of the TM in the image encoder MMP. Experimental results are discussed in Section IV and conclusions are presented in Section V.

This project was funded by FCT - "Fundação para a Ciência e Tecnologia", Portugal, under the grant SFRH/BD/45460/2008, and Project COMUVI (PTDC/EEA-TEL/099387/2008)

II. Template matching prediction

TM was initially proposed as a texture synthesis tool [4]. Nevertheless, promising results have been reported by employing texture synthesis algorithms in spatial and temporal prediction, namely intra [5] and inter [6] frame prediction for image and video coding.

The major advantage of TM is the prediction of the target block without transmitting any additional displacement information. Since the TM prediction step is performed using a causal search area of the reference image, the same predicted block can be determined both at the encoder and at the decoder.

The TM performance relies on the correlation between pixels of the target block and the reconstructed pixels surrounding it. The template area is composed by reconstructed pixels of the causal neighborhood, usually belonging to the left and top margins of the target block. The prediction of the target block is determined by minimizing the matching error between its template and each template lying in the search area. The TM prediction is expected to have a good performance for areas of the image with high correlation between the target block and its template.

In this paper we optimized TM for stereo images' prediction. Instead of the traditional use of a causal search region inside the current image, the matching of the template region is done in the reference image. The dimensions of TM search area were also adjusted considering the features of stereo disparity. The maximum height of the TM search area is set to a small value, while the maximum horizontal shift is dynamically determined for each stereo pair.

In order to determine the maximum horizontal disparity for each stereo pair, we propose a fixed block size matching procedure along the horizontal direction. A cumulative histogram of displacements is then computed for all non-overlapped blocks of the image. The maximum allowed disparity in each direction is determined in the cumulative histogram, by computing the shift that comprises the cumulative probability of 90% plus an offset of three pixels. This is performed both for positive and negative horizontal displacements. The computed bounds are transmitted to the decoder on the image's header and used in the matching stage of the algorithm. This also has a useful side-effect of reducing the computational complexity of the algorithm.

Since stereo disparities correspond mostly to horizontal displacements, namely in aligned stereo pairs, the vertical length of TM search area was reduced. A fixed maximum template displacement of 3 pixels above and 3 pixels below the position of the target block was set. Experimental results have shown that the chosen vertical bounds are a good compromise between computational complexity and coding performance for a wide set of used test images.

The block matching efficiency of the TM prediction was also improved by using sub pixel accuracy. For this purpose, a quarter-pixel precision was used in the search for the best matched template. Interpolated pixels are determined using the well known method defined in the H.264/AVC standard.

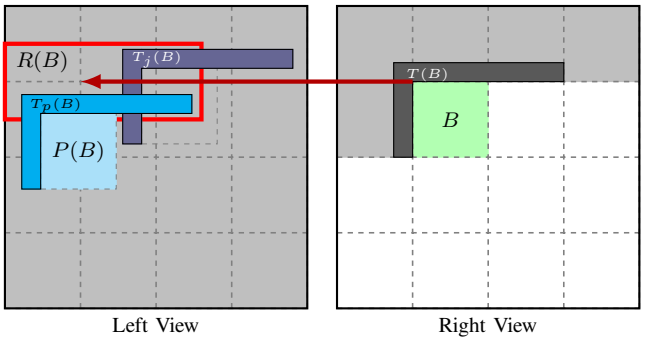


Fig. 1: Static template matching prediction.

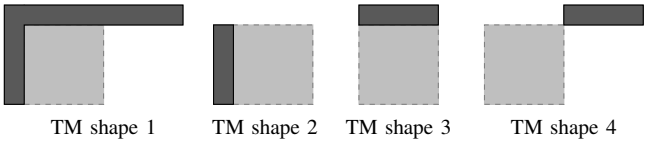


Fig. 2: Dynamic template matching prediction.

The template's shape used for the proposed prediction was also optimized for stereo images. To better exploit the TM, two different approaches for TM-based stereo prediction were studied, the static and dynamic TM. Proposed TM implementations are presented in the following sub-sections.

A. Static template matching

Static TM refers to the common case of the TM algorithm, that uses a fixed template to find the best predictor for the target block. Figure 1 represents a diagram of static TM applied to stereo prediction. The proposed template for static TM, denoted by $T(B)$, with B being the target block, is represented by the dark pixels in Figure 1. The horizontal length of the top pixels of template includes the top-right block. However, the right portion of template top pixels should be ignored when these pixels are not available.

The goal of TM is to solve the minimization problem of distance d_j between the template, $T(B)$, and any candidate template, $T_j(B)$, belonging to the search area, $R(B)$. The proposed minimization problem is given as

$$p = \underset{j \in [1 \dots M]}{\arg\min} \{d_j : d_j = SSD\,(T(B), T_j(B))\} \qquad (1)$$

where SSD stands for the sum of squared differences. Other distance metric could be chosen for this minimization problem. The candidate templates, $T_j(B)$, are extracted from the search area, $R(B)$, located in the reconstructed left image, with M possible templates.

In stereo images it is expected that the displacement between $T(B)$ and $Tp(B)$ matches the disparity of the templates. Since TM assumes that the template has high correlation with the target block, the predictor of the target block B is given by the block $P(B)$, adjacent to the best-matched template $Tp(B)$.

B. Dynamic template matching

Dynamic TM [7] is an improvement to the TM prediction algorithm. In dynamic TM various templates, with different shapes, are used to solve the minimization problem presented

in (1). The optimum template shape is selected and a symbol, identifying the chosen template, is encoded and transmitted to the decoder.

Figure 2 illustrates the four chosen template shapes of the proposed dynamic TM, that were found to be suited for stereo image coding, after a series of experimental tests. For each template shape, the static TM algorithm is performed. The dynamic TM selects the best template shape that minimizes the following Lagrangian cost function

$$J_k = SSD(B, P_k(B)) + \lambda R(k) \ , \quad k \in [1 \dots 4], \qquad (2)$$

where $R(k)$ is the rate needed to encode the chosen template shape k. The SSD is computed between the block $P_k(B)$ predicted with the template shape k and the target block B.

Contrary to the static TM, the dynamic approach requires some additional overhead to indicate the selected shape. However, as presented in Section IV, our experimental results show that this strategy is more efficient than the static TM algorithm.

III. Stereo Image Encoder with Dynamic Template Matching

In order to encode stereo images, we propose an encoder that combines the proposed prediction mode with the MMP algorithm [2]. The MMP algorithm is a generic lossy data compression method that has been successfully applied to grayscale image coding. MMP divides the image into non-overlapping blocks and uses patterns at different scales from an adaptive dictionary to approximate them. Using TM with the MMP, the proposed scheme for stereo image coding comprises a fully pattern matching based encoder. The MMP algorithm uses an hierarchical prediction scheme, based in the same prediction modes as the H.264/AVC standard, plus the LSP (least squares prediction) mode [8]. The residue is encoded via pattern matching. MMP uses an adaptive block size for prediction, with blocks that range from 4×4 to 16×16. This flexible partitioning of the prediction step, and the partitioning of residue blocks until 1×1, is an important feature of the MMP algorithm [3].

When a stereo pair is being encoded, the left image is encoded with the original MMP algorithm, that uses the basic set of intra prediction modes. To encode the right image, the TM prediction mode was used as an additional mode, for the MMP image encoder. Figure 3 shows a typical usage rate of the used prediction modes, for the compression of the right view of the Tsukuba stereo pair, at 0.3 bpp. As can be seen, the TM prediction mode is more efficient, and used more often, than intra prediction modes because it exploits the high redundancy between both views.

To signal the use of the TM mode, an additional symbol was created, denoted inter-flag. The inter-flag informs the decoder whether the mode is the TM mode or one of the original intra modes. The flag is encoded with adaptive arithmetic encoder using two contexts. In order to select the context, the prediction mode used in the top and left neighbors blocks is inspected. If both blocks were predicted with TM, a context is chosen, otherwise another context is selected.

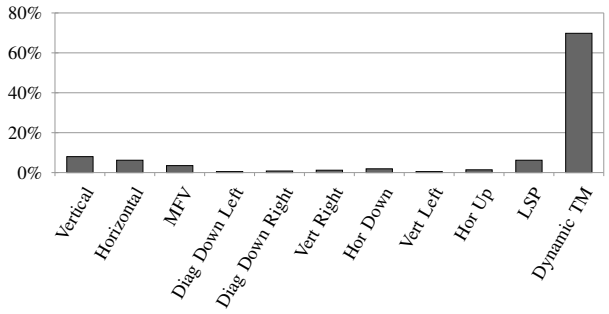
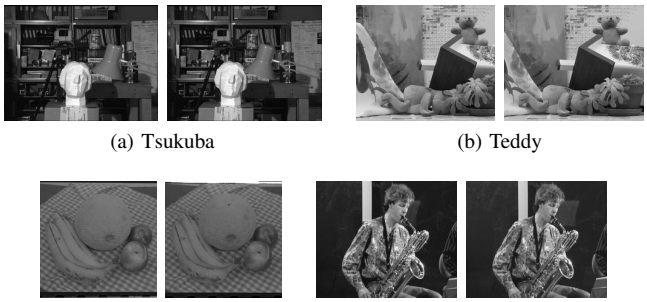


Fig. 3: Usage rate of each mode for right image of Tsukuba.



(a) Tsukuba (b) Teddy

(c) Fruit (d) Saxo

Fig. 4: Test images.

The MMP flexible segmentation is one of its advantageous features, however it has some impact in the TM prediction performance. Larger blocks tend to be used in areas where the disparity varies smoothly, while smaller ones are used in areas where there is a higher disparity variation. When using larger blocks, e.g. 16×16 blocks, TM uses a neighborhood with 4 pixel lines. Since smaller blocks are used in areas with higher disparity variation, this large neighborhood might contain pixels with disparity values that are not correlated with the block disparity values. To improve TM performance, a smaller neighborhood is used for smaller blocks. In the proposed algorithm, the number of rows/columns used for the TM neighborhood, i.e. the template thickness th, is determined by $th = (B_x + B_y)/8$, where B_x and B_y correspond to the block width and height, respectively.

IV. Experimental Results

In order to evaluate the rate-distortion performance of the proposed prediction mode, we compared it with the state-of-the-art stereo image encoder H.264/AVC Stereo Profile [9]. The software used in our simulations is JM-17.2, using the Stereo High Profile.

The test set comprises the stereo images shown in Figure 4. Most of these images were obtained with a parallel camera arrangement, such as the Tsukuba and Teddy. Nevertheless, we have also included stereo images, such as Fruit and Saxo, that were captured with convergent cameras to be more thorough.

Figures 5-8 show the rate-distortion curves for both views of each stereo pair. Only right image benefits from the use of the TM prediction mode. The results for both static and dynamic TM algorithms are also shown. As can be seen, in all cases the dynamic TM prediction mode presents a consistent gain over

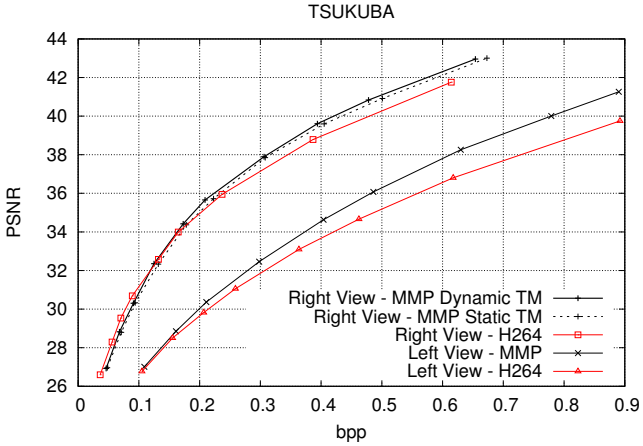


Fig. 5: Experimental results for stereo image Tsukuba.

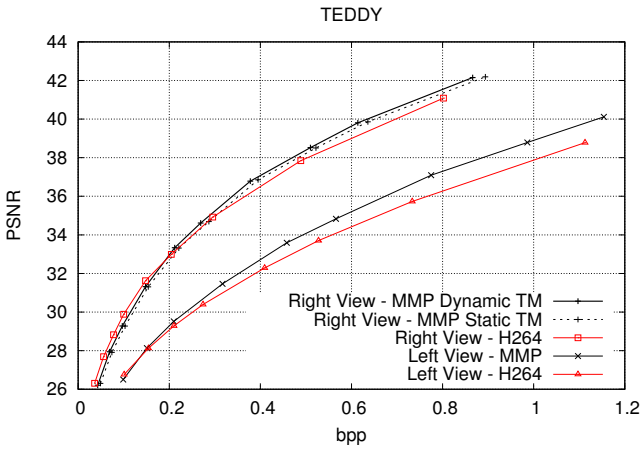


Fig. 6: Experimental results for stereo image Teddy.

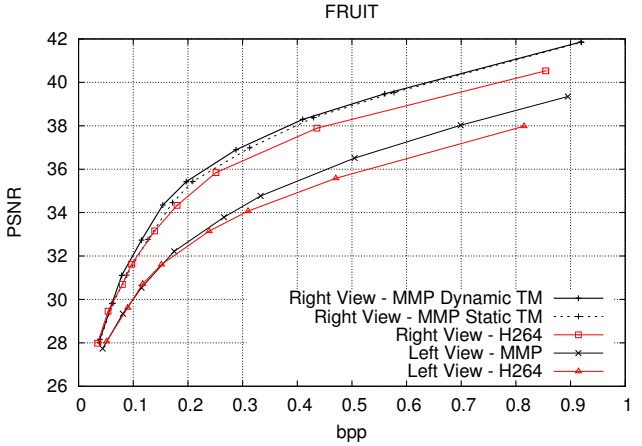


Fig. 7: Experimental results for stereo image Fruit.

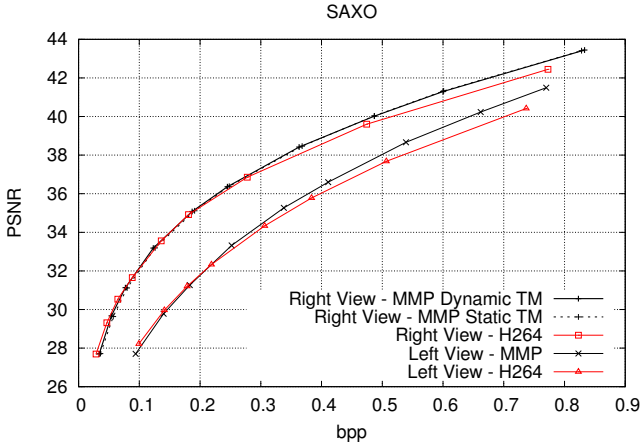


Fig. 8: Experimental results for stereo image Saxo.

the static TM prediction mode, for all compression rates. This results from the fact that dynamic TM may choose a different neighborhood, that can be more correlated to the target block disparity than the neighborhood for static TM.

Figures 6 and 8 present a rate-distortion gain up to 0.5 dB over H.264/AVC standard for the right image. Gains of almost 1 dB at higher rates can be seen at Figures 5 and 7. One may notice that achieved gain may vary with input stereo images. Despite some losses in some cases at low rates, the proposed scheme is consistently more efficient than H.264/AVC standard at medium and high rates. When compared with MMP without the use of TM, note that the RD performance for this case, is generally similar to that achieved for the left image. These results show the efficiency of the TM algorithm in finding a good prediction match.

All figures show a superior rate-distortion performance of the MMP for the reference (left) images.

V. Conclusion

In this paper we propose a TM based inter prediction method for stereo images. The proposed prediction mode was combined with the MMP algorithm to encode images with a fully pattern matching based scheme.

Our experimental results show that the right view can be efficiently predicted with template matching, namely dynamic TM, which is able to achieve gains over the H.264/AVC standard in most cases. In future work, the use of different disparity estimation algorithms will be investigated.

References

[1] M. Lukacs, "Predictive coding of multi-viewpoint image sets," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, vol. 11, Apr. 1986, pp. 521 – 524.

[2] N. Rodrigues, E. da Silva, M.de Carvalho, S. de Faria, and V. Silva, "On dictionary adaptation for recurrent pattern image coding," *Image Processing, IEEE Transactions on*, vol. 17, no. 9, pp. 1640–1653, September 2008.

[3] N. Francisco, N. Rodrigues, E. da Silva, M. de Carvalho, S. de Faria, and V. Silva, "Scanned compound document encoding using multiscale recurrent patterns," *Image Processing, IEEE Transactions on*, vol. 19, no. 10, pp. 2712–2724, October 2010.

[4] L.-Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," *Proceeding of SIG-GRAPH*, pp. 479–488, July 2000.

[5] T. Tan, C. Boon, and Y. Suzuki, "Intra prediction by template matching," *IEEE Int. Conf. on Image Proc.*, October 2006.

[6] K. Sugimoto, M. Kobayashi, Y. Suzuki, S. Kato, and C. S. Boon, "Inter frame coding with template matching spatio-temporal prediction," *IEEE Int. Conf. on Image Proc.*, October 2004.

[7] M. Turkan and C. Guillemot, "Image prediction: Template matching vs. sparse approximation," *IEEE Int. Conf. on Image Proc.*, September 2010.

[8] D. B. Graziosi, N. Rodrigues, E. da Silva, S. de Faria, and M. de Carvalho, "Improving multiscale recurrent pattern image coding with least-squares prediction," in *IEEE Int. Conf. on Image Proc.*, Nov 2009.

[9] http://iphome.hhi.de/suehring/tml/download/.