Hybrid template and block matching algorithm for image intra prediction

Safa Cherigui, Christine Guillemot, Dominique Thoreau, Philippe Guillotel, Patrick Pérez

HYBRID TEMPLATE AND BLOCK MATCHING ALGORITHM FOR IMAGE INTRA PREDICTION

Safa Cherigui [1,2], Christine Guillemot [1], Dominique Thoreau [2], Philippe Guillotel [2] and Patrick Perez [2]

[1] *INRIA*
Campus de Beaulieu, 35042 Rennes Cedex France
firstname.lastname@irisa.fr

[2] *Technicolor Research and Innovation*
Av. Belle Fontaine, 35576 Cesson-Sévigné France
firstname.lastname@technicolor.com

ABSTRACT

Template matching has been shown to outperform the H.264 prediction modes for Intra video coding thanks to better spatial prediction and no additional ancillary data to transmit. The method indeed works well when the template and the block to be predicted are highly correlated, e.g., in homogenous image areas, however, it obviously fails in areas with non homogeneous textures. This paper explores the idea of using a block-matching intra prediction algorithm which, thanks to a Rate-Distorsion (RD) based decision mechanism, will naturally be used in image areas when template matching (TM) fails. This new method offers a significant coding gain compared to H.264 Intra prediction modes and the template matching based prediction. Indeed, the TM-based algorithm and the proposed hybrid algorithm lead, with the Bjontergaard measure, to rate gains of up to respectively 38.02% and 48.38% at low bitrates when compared with H.264 Intra only.

Index Terms: Texture prediction, intra coding, template matching, block matching, H.264.

1. INTRODUCTION

This paper addresses the problem of Intra prediction which is a key component of image and video compression algorithms. Given observations, or known samples in a spatial neighborhood which may be affected by quantization noise, the goal is to estimate unknown samples of the block to be predicted. For example, in H.264/AVC, there are three intra-frame prediction types called Intra-4x4, Intra-8x8 and Intra-16x16 respectively [1]. For each intra prediction mode, a block is predicted from prior encoded samples of spatially neighboring blocks. In addition to the so-called "DC" mode which consists in predicting the entire block from the mean of neighboring pixels, eight directional prediction modes have been specified. The prediction is done by simply "propagating" the pixel values along the specified direction. This approach is particularly suitable in presence of contours, when the directional mode chosen corresponds to the orientation of the contour. However, it fails in more complex textured areas. Besides, in H.264/AVC, intra prediction is used to eliminate the correlation between current block and adjacent pixels, but for some video sequences, the current block has a great similarity with blocks from other parts of the picture.

Several new intra prediction methods have been proposed to exploit redundancy within an image. Firstly, a block based intra prediction method has been introduced in [2], which integrates motion estimation and motion compensation techniques classically used in inter-frame. In this approach, the reconstructed part of the image is taken as a reference frame, and the prediction is the block that best matches the current block. Note that the derived matching vectors are required to be coded because they are needed at the decoder for reconstruction of the prediction. An alternative spatial prediction algorithm based on template matching has been described in [3]. A so-called template is formed by previously encoded pixels in the close neighborhood of the block to be predicted. The best match between the template of the block to be predicted and candidate texture patches with the same shape, within a causal search window, gives the predictor for the current block to be coded. This approach has later been improved in [4] by averaging multiple template matching predictors, including larger and directional templates, resulting in up to 15% rate saving when compared to H.264/AVC Intra-prediction. However, a best match of the template does not necessarily lead to a good predictor for the unknown block to be predicted, especially in non homogenous texture areas of the image or when there are discontinuities between the template and the block to be predicted. Blocks located in image areas having such characteristics will be better predicted if we can do a direct match with the block to be predicted. But this is feasible only at an extra rate cost which will represent the coding of the vector of correspondence between the two matching blocks. This is a block matching algorithm as used classically for motion estimation, but this time run spatially by defining a search window in the causal part of the current frame. In order to avoid the strong impact of the matching vector cost on the rate-distortion performance of the complete compression algorithm, this mode has to be used only for blocks where TM fails. A decision based on an RD cost function including the rate of the matching vector

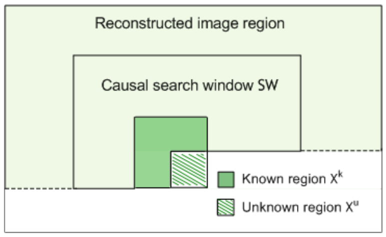

Fig. 1. Prediction problem statement and notations: X^k denotes the known pixels (or template), X^u is the current block to be predicted, and SW is the window within which the best predictor will be searched for.

and the prediction residue is thus proposed. More precisely, two H.264 intra prediction modes (the least statistically used modes among the 9 available ones) are replaced by the TM and BM based intra predictions. The TM prediction algorithm and the proposed hybrid TM/BM predictions algorithm can thus be compared with the H.264 intra prediction modes.

The rest of the paper is organized as follows. Section 2 formulates the image prediction problem and reviews the main background approaches. Section 3 describes the proposed hybrid template and block matching intra predictions algorithm. Section 4 gives performance illustrations in the context of prediction and compression.

2. IMAGE PREDICTION: BACKGROUND APPROACHES

Let X be a texture patch which comprises a known part X^k (of a given shape) formed by the pixels located in a causal neighborhood of the current block, called the template, and of an unknown part X^u formed by the current block to be predicted (see Fig.1). This section briefly reviews the most popular approaches to predict X^u knowing the previously encoded and decoded samples of X^k.

2.1. H.264 prediction modes

The H.264/AVC standard defines three types of intra prediction modes depending on the block size: the Intra-4x4, Intra-8x8 and Intra-16x16 [1]. For Intra prediction of 16x16 blocks, four intra prediction modes can be used while for Intra prediction of 4x4 and 8x8 blocks, nine modes have been defined as shown in Fig. 2. Here, we consider the three types of blocks which are then predicted from prior encoded pixels from spatially neighboring blocks. In addition to the so-called "DC" mode which consists in predicting the entire 4x4 or 8x8 block from the mean of neighboring pixels, eight directional prediction modes are specified [5]. The prediction is done by simply propagating (or interpolating) the pixel values along the specified direction.

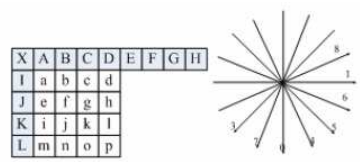

Fig. 2. A 4x4 block with its boundary pixels and its eight directional intra prediction modes.

2.2. Template Matching (TM)

The TM algorithm searches, within a given search window SW located in the causal part of the image, the "best matching" patch in the sense of minimizing the Euclidean distance between the vector formed by the known samples X^k of the neighborhood of the block to be predicted (of n^2 pixels size) and the co-located pixels in the candidate patches taken from SW. In Fig.1, the known region X^k is formed by the pixel values of 3 neighboring blocks, hence, the corresponding X^k vector is of size $N = 3n^2$ pixels. Note, however, that different forms of templates X^k can be considered [6]. The pixels in the candidate patch which are co-located to those of the block to be predicted are then used as predictors.

2.3. Block Matching (BM)

The block matching algorithm has been initially used for motion compensation in video compression. However, this approach has also been explored for intra coding [2]. This block based intra prediction method performs similarly as motion estimation and compensation in inter-frame predictive coding, but here the reference block used in the prediction is searched in the causal part of the current image. Indeed, this algorithm searches, in a given search window SW in the reconstructed image, the best matching block by minimizing the Euclidian distance between the vector formed by the unknown samples X^u of the block to be predicted and the candidate blocks taken from SW. Note that the unknown region X^u is of size n^2 pixels.

3. HYBRID TM/BM PREDICTIONS ALGORITHM

The proposed intra prediction method consists in using the best of the TM and BM algorithms. The BM spatial prediction is used when the TM spatial prediction fails. Using both TM and BM intra prediction modes enables to take advantage of both methods. Indeed, the template matching approach works well when the unknown block to be predicted and its template are correlated to the ones matched by template. Besides, TM does not require sending additional matching vectors but the same template matching process has to be carried out on the decoder side increasing the complexity. By contrast, block matching based prediction method can better

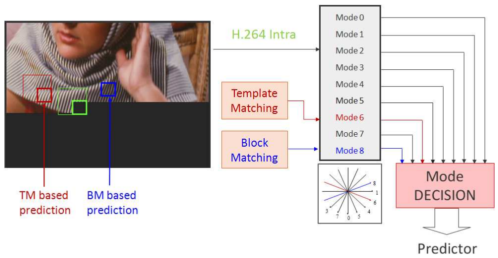

Fig. 3. Synoptic of hybrid TM/BM intra prediction modes selection for 4x4 or 8x8 blocks.

decorrelate the block to be predicted with the reconstructed image region and does not require to perform any estimation motion on the decoder side but the matching vectors have to be sent to the decoder. Therefore, the hybrid TM/BM method can first reduce the computational complexity at the decoder compared to TM only and also reduce the impact of matching vectors rate cost compared to BM only.

We have introduced TM and BM-based intra prediction modes in the H.264 codec (KTA reference software [7]) by replacing two existing modes of the H.264 algorithm as depicted in Fig.3. Here, the two intra prediction modes, which are replaced by TM and BM modes, correspond to the least statistically used modes in H.264 Intra only. This facilitates the implementation by keeping the same syntax and allows saving signaling bits. Here, the proposed hybrid TM/BM approach is applied to Intra-4x4 and Intra-8x8 and the matching vectors are limited to full-pel accuracy. The selection of the best coding mode for a given block is conducted according to a minimization of Lagrangian criterion, also called Rate-Distortion Optimization (RDO) criterion:

$$\min \left(D + \lambda \times R \right) \tag{1}$$

where D represents the distortion between the original block and the reconstructed block by using the Sum of Square Errors (SSE) distance metric, λ represents a trade-off coefficient between the distortion and the bits needed for coding the block. The parameter λ is selected as in the KTA reference software ($\lambda = 0.65 \times 2^{QP/3}$), with QP corresponding to the quantization parameter. For intra coding mode based on BM, the rate cost R is calculated according to the following formula:

$$R = R_{mode} + R_{res} + R_{vector} \tag{2}$$

where R_{mode}, R_{res} and R_{vector} represent the bits needed for coding respectively the block coding mode, the residue and the matching vector. R_{mode} and R_{res} are encoded with the KTA entropy coding scheme (CABAC). The matching vectors are coded using a fixed length code (FLC) which varies nevertheless according to the size of the causal search window SW. For intra coding mode based on TM, the equation (2) is also used with $R_{vector} = 0$.

4. PERFORMANCE ILLUSTRATION

Three coding schemes: the TM-based algorithm, the BM-based algorithm and the proposed hybrid TM/BM based algorithm have been assessed comparatively to H.264 Intra prediction modes. Several test sequences of different resolutions were tested and shown in Tables 1 and 2, where coding efficiencies were compared using the Bjontegaard metric [8]. Here, we considered a search window of size of 64x128 pixels and matching vectors at full-pel accuracy. Compared to H.264 Intra only, average BD-rate gains obtained with the TM-based compression method are of 17.56% at low bitrates and 13.64% at high bitrates whereas the ones obtained with the hybrid TM/BM based compression method are of 24.78% at low bitrates and 16.74% at high bitrates.

The first set of pictures (from Barbara to City2) is partially composed of pseudo-regular textures that fit well with the different algorithms tested. The pictures Zone1 (Fresnel image) and Pan0_qcif (nearly regular texture) can be considered as atypical. The Matrix, composed of sequences of random numbers, is a test image (taken from the Matrix movie) that highlights the virtues of two algorithms. Indeed, the TM does not allow a reliable prediction of a given number knowing only its neighborhood. However, this approach works well in texture synthesis because it is based only on visual aspects instead of a PSNR measure. In contrast, the BM allows the good prediction of a number, if the same number is present in the causal search window. We obtained also interesting results with conventional images like Foreman (CIF) and Spincalendar (HD), especially at low bitrates.

From all test sequences, we can notice that the use of both TM and BM predictions enables to get better coding efficiency compared to H.264 Intra for various contents in all resolutions. However, the proposed hybrid TM/BM method offers variable coding performances due to the type of content in the pictures. In fact, on the one hand, we can see that the coding gains are less important for the sequences Barbara and Foreman compared to H.264 Intra only since H.264 intra prediction modes are quite well suited. On the other hand, we can observe that the coding gains are significant for the pictures such as Zone1 and City2 that present local stationary properties. Indeed, a bit rate saving up to 41.78% and 48.38% respectively is achieved at low-bitrates compared to H.264 Intra only.

Experimental results show also that the proposed method outperforms the TM-based approach and the BM-based approach. This was explained by the fact that the two approaches are quite complementary. From the simulation results, we can point out that BM provides a better prediction in local variations of the signal than TM as we can see with the pictures Barbara and Snook (cf. Table 1).

Fig. 4 illustrates the use of TM and BM prediction modes for Intra-8x8 only. We can notice that the picture Barbara presents pseudo-periodic structures which are locally vari-

Low bit rates	TM		BM		TM + BM	
QP:16,21,26,31	dB	% rate	dB	% rate	dB	% rate
Barbara	0.45	-6.80	0.49	-7.41	0.83	-12.06
Wool	0.88	-13.81	1.13	-17.36	1.37	-20.93
Snook	1.08	-14.77	1.47	-20.25	1.58	-21.71
Zone1	5.87	-38.02	5.98	-38.76	6.60	-41.78
Pan0_qcif	1.21	-18.91	1.21	-19.16	1.64	-25.36
City2	3.32	-37.92	4.03	-44.21	4.42	-48.38
Matrix	0.85	-9.97	1.18	-15.33	1.78	-21.33
Spincalendar	0.56	-10.53	0.86	-15.74	1.13	-20.29
Foreman	0.38	-7.34	0.40	-7.67	0.58	-11.21
Average	**1.62**	**-17.56**	**1.86**	**-20.65**	**2.21**	**-24.78**

Table 1. Rate-Distortion gains (with the Bjontergaard measure) with respect to H.264 Intra prediction modes at low bit rates (computed at 4 rate points: QP=16,21,26,31).

High bit rates	TM		BM		TM + BM	
QP:26,31,36,41	dB	% rate	dB	% rate	dB	% rate
Barbara	0.45	-5.41	0.36	-4.38	0.61	-7.37
Wool	0.79	-8.57	0.75	-8.46	0.95	-10.53
Snook	1.31	-11.20	1.53	-13.13	1.65	-14.12
Zone1	6.47	-35.89	7.35	-38.83	7.66	-40.42
Pan0_qcif	1.44	-12.96	1.06	-10.00	1.60	-14.48
City2	4.66	-32.14	4.62	-33.05	5.20	-36.56
Matrix	0.75	-6.16	0.90	-7.52	1.42	-11.84
Spincalendar	0.56	-6.89	0.59	-7.52	0.86	-10.73
Foreman	0.25	-3.57	0.22	-3.31	0.32	-4.62
Average	**1.85**	**-13.64**	**1.93**	**-14.02**	**2.25**	**-16.74**

Table 2. Rate-Distortion gains (with the Bjontergaard measure) with respect to H.264 Intra prediction modes at high bit rates (computed at 4 rate points: QP=26,31,36,41).

ables due to spatial deformations of the fabric. It is the reason why the TM mode is less selected than BM mode. The picture City2 shows that TM and BM methods complement each other since this image presents locally homogeneous textures while at the same time having a slight perspective effect.

Note that at the encoder, the BM algorithm can be performed jointly with the TM algorithm. However, at the decoder, the computational complexity depends only on the occurence of TM mode.

5. CONCLUSION

In this paper, a hybrid template and block matching algorithm for intra coding has been introduced. This new approach of prediction offers interesting results compared to respectively H.264 Intra only and TM mode included in H.264 Intra. Moreover, the hybrid TM/BM method enables to reduce

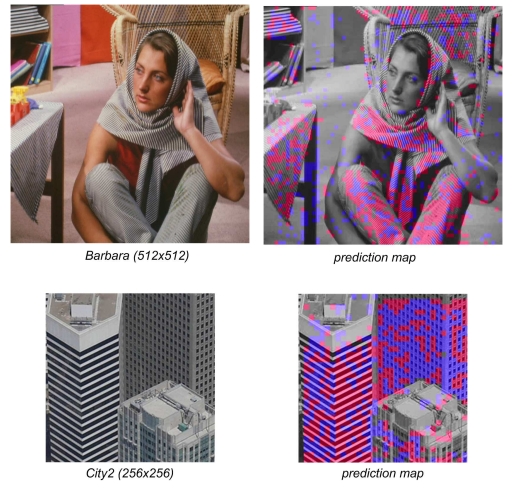

Fig. 4. Illustration of modes used in 8x8 prediction: TM (in blue) and BM (in red) (QP=31).

the computational complexity at the decoder compared to TM only because the BM mode does not need to perform a best matching search at the decoder. Future work will consider more efficient methods using a vector predictor to code the BM vectors such as a median predictor. Besides, so far, TM and BM have been performed only at full-pel accuracy, which enables us to avoid sub-pel interpolations on the fly. However, the sub-pel accuracy could be also investigated knowing that the sub-pel accuracy will increase the BM vectors cost.

6. REFERENCES

[1] T. Wiegand, G. J. Sullivan, B. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13-7, pp. 560–576, Jul. 2003.

[2] S.-L. Yu and C. Chrysafis, “New intra prediction using intra macroblock motion compensation”, in *JVT-C151, 3rd meeting of Joint Video Team (JVT))*, May 2002.

[3] T. K. Tan, C. S. Boon, and Y. Suzuki, “Intra prediction by template matching”, in *IEEE Int. Conf. Image Process. (ICIP)*, 2006, pp. 1693–1696.

[4] T. K. Tan, C. S. Boon, and Y. Suzuki, “Intra prediction by averaged template matching predictors”, in *IEEE Conf. Consumer Comm. Network. Conf. (CCNC)*, 2007, pp. 405–409.

[5] “Draft ITU-T Rec. and Final Draft Int. Standard of Joint Video Spec.”, ITU-T Rec. H.264. ISO/IEC 14496-10 AVC, 2003.

[6] M. Turkan and C. Guillemot, “Image prediction: Template matching vs. Sparse approximation”, in *IEEE Int. Conf. Image Process. (ICIP)*, 2010, pp. 789–792.

[7] “http://iphome.hhi.de/suehring/tml/download/kta/”.

[8] G. Bjontegaard, “Calculation of average psnr differences between rd curves”, in *document VCEG-M33, ITU-T VCEG Meeting*, April 2001.