Hybrid template and block matching algorithm for image intra prediction
Safa Cherigui, Christine Guillemot, Dominique Thoreau, Philippe Guillotel, Patrick Pérez

To cite this version:
Safa Cherigui, Christine Guillemot, Dominique Thoreau, Philippe Guillotel, Patrick Pérez. Hybrid template and block matching algorithm for image intra prediction. ICASSP’12 - IEEE International Conference on Acoustics, Speech, and Signal Processing, Mar 2012, Kyoto, Japan. pp.4. hal-00695277

HAL Id: hal-00695277
https://hal.archives-ouvertes.fr/hal-00695277
Submitted on 8 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
HYBRID TEMPLATE AND BLOCK MATCHING ALGORITHM FOR IMAGE INTRA PREDICTION

Safa Cherigu1,2, Christine Guillemot1, Dominique Thoreau2, Philippe Guillotel2 and Patrick Perez2

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 homogeneous 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-Distortion (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 Bjontegaard 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...
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
where \( R \) is also used with dow nevertheless according to the size of the causal search win-
tors are coded using a fixed length code (FLC) which varies KT A entropy coding scheme (CABAC). The matching vec-
formula: BM, the rate cost the quantization parameter. For intra coding mode based on
coding respectively the block coding mode, the residue and 
ors (SSE) distance metric, 
where \( D \) represents the distortion between the original block 
and the reconstructed block by using the Sum of Square Er-
errors (SSE) distance metric, \( \lambda \) represents a trade-off coefficient 
between the distortion and the bits needed for coding 
the block. The parameter \( \lambda \) is selected as in the KTA refer-
cence 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:
\[
R = R_{\text{mode}} + R_{\text{res}} + R_{\text{vector}}
\]  
where \( R_{\text{mode}}, R_{\text{res}} \) and \( R_{\text{vector}} \) represent the bits needed for 
coding respectively the block coding mode, the residue and 
the matching vector. \( R_{\text{mode}} \) and \( R_{\text{res}} \) are encoded with the 
KTA entropy coding scheme (CABAC). The matching vec-
tors are coded using a fixed length code (FLC) which varies 
nevertheless according to the size of the causal search win-
dow \( SW \). For intra coding mode based on TM, the equation 
(2) is also used with \( R_{\text{vector}} = 0 \).

4. PERFORMANCE ILLUSTRATION

Three coding schemes: the TM-based algorithm, the BM-
based algorithm and the proposed hybrid TM/BM based al-
gorithm have been assessed comparatively to H.264 Intra pre-
diction modes. Several test sequences of different resolutions 
were tested and shown in Tables 1 and 2, where coding ef-
ciciencies were compared using the Bjontegaard metric [8]. Here, we considered a search window of size of 64x128 pix-
els 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 par-
tially composed of pseudo-regular textures that fit well with 
the different algorithms tested. The pictures Zone1 (Fresnel image) and Pan0_qcif (nearby regular texture) can be consid-
ered as atypical. The Matrix, composed of sequences of ran-
dom 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 know-
ing 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 al-

Experimental results show also that the proposed method 
outperforms the TM-based approach and the BM-based ap-
proach. This was explained by the fact that the two ap-
proaches 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-

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

<table>
<thead>
<tr>
<th></th>
<th>TM</th>
<th>BM</th>
<th>TM + BM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbara</td>
<td>0.45 -6.80</td>
<td>0.49 -7.41</td>
<td>0.83 -12.06</td>
</tr>
<tr>
<td>Wool</td>
<td>0.88 -13.81</td>
<td>1.13 -17.36</td>
<td>1.37 -20.93</td>
</tr>
<tr>
<td>Snook</td>
<td>1.08 -14.77</td>
<td>1.47 -20.25</td>
<td>1.58 -21.71</td>
</tr>
<tr>
<td>Zone1</td>
<td>5.87 -38.02</td>
<td>5.98 -38.76</td>
<td>6.60 -41.78</td>
</tr>
<tr>
<td>Pan0_qcif</td>
<td>1.21 -18.91</td>
<td>1.21 -19.16</td>
<td>1.64 -25.36</td>
</tr>
<tr>
<td>City2</td>
<td>3.32 -37.92</td>
<td>4.03 -44.21</td>
<td>4.42 -48.38</td>
</tr>
<tr>
<td>Matrix</td>
<td>0.85 -9.97</td>
<td>1.18 -15.33</td>
<td>1.78 -21.33</td>
</tr>
<tr>
<td>Spincalendar</td>
<td>0.56 -10.53</td>
<td>0.86 -15.74</td>
<td>1.13 -20.29</td>
</tr>
<tr>
<td>Foreman</td>
<td>0.38 -7.34</td>
<td>0.40 -7.67</td>
<td>0.58 -11.21</td>
</tr>
<tr>
<td>Average</td>
<td>1.62 -17.56</td>
<td>1.86 -20.65</td>
<td>2.21 -24.78</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>TM</th>
<th>BM</th>
<th>TM + BM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbara</td>
<td>0.45 -5.41</td>
<td>0.36 -4.38</td>
<td>0.61 -7.37</td>
</tr>
<tr>
<td>Wool</td>
<td>0.79 -8.57</td>
<td>0.75 -8.46</td>
<td>0.95 -10.53</td>
</tr>
<tr>
<td>Snook</td>
<td>1.31 -11.20</td>
<td>1.53 -13.13</td>
<td>1.65 -14.12</td>
</tr>
<tr>
<td>Zone1</td>
<td>6.47 -35.89</td>
<td>7.35 -38.83</td>
<td>7.66 -40.42</td>
</tr>
<tr>
<td>Pan0_qcif</td>
<td>1.44 -12.96</td>
<td>1.06 -10.00</td>
<td>1.60 -14.48</td>
</tr>
<tr>
<td>City2</td>
<td>4.66 -32.14</td>
<td>4.62 -33.05</td>
<td>5.20 -36.56</td>
</tr>
<tr>
<td>Matrix</td>
<td>0.75 -6.16</td>
<td>0.90 -7.52</td>
<td>1.42 -11.84</td>
</tr>
<tr>
<td>Spincalendar</td>
<td>0.56 -6.89</td>
<td>0.59 -7.52</td>
<td>0.86 -10.73</td>
</tr>
<tr>
<td>Foreman</td>
<td>0.25 -3.57</td>
<td>0.22 -3.31</td>
<td>0.32 -4.62</td>
</tr>
<tr>
<td>Average</td>
<td>1.85 -13.64</td>
<td>1.93 -14.02</td>
<td>2.25 -16.74</td>
</tr>
</tbody>
</table>

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.

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

6. REFERENCES