INTRA-FRAME PREDICTION WITH LAPPED TRANSFORMS FOR IMAGE CODING

Rafael Galvão de Oliveira and Béatrice Pesquet-Popescu

Département Traitement du Signal et des Images, Télécom ParisTech
46 rue Barrault, 75634 Paris, France

ABSTRACT
In this paper we propose the use of intra-frame prediction with lapped transforms for image coding. Both lapped transforms and intra prediction exploit the redundancies of neighboring blocks and the combination of the two techniques results in a very efficient image coding scheme. The difficulty to combine them comes from the necessity to use for the prediction of the current block pixels in the causal neighborhood that have not been completely processed by the overlapping transform. We show how to overcome this difficulty, and thus the system presented here outperforms the traditional one, using intra-prediction and DCT, and the direct application of lapped transforms in all tested images.

Index Terms— Lapped transforms, intra-frame prediction, image coding, H.264/AVC

1. INTRODUCTION
One of the main motivations for using transforms with overlap in coding schemes is trying to improve objective and subjective performance when compared with the block transforms. Since not only the samples of the block are used in the transform, it is possible to exploit similarities present in the neighboring blocks. Other very positive consequence of the overlapping is the reduction of the blocking effect, i.e. discontinuities in the block borders caused by the quantization of the transform coefficients.

Since the early development of lapped transforms [1], they have been compared to DCT and wavelets. In [2], there have been many comparisons among transforms such as the 9/7-biorthogonal wavelet transforms used in JPEG2000 [3], lapped orthogonal transforms (LOT) [4], generalized LOT (GenLOT) [5] and generalized lapped biorthogonal transforms (GLBT) [6]. The comparisons were also carried out with JPEG [7], SPIHT [8,9] and JPEG2000 [10]. The results have shown consistent improvement in all tested coders. It has also been shown that, in some cases, the lapped transforms can also outperform the scheme of intra-frame prediction with DCT used in the H.264/AVC [11,12], which was shown to have comparable or better performance for still image coding than JPEG2000[13].

More recently, a new still-image coding standard, JPEG XR [14] (lately known as Windows Media Photo and HD Photo), has been adopted and proposes an optional additional overlapping core transform, which consists in a reversible lapped biorthogonal transform. It is, up to this moment, the only image coding standard that uses lapped transforms. Contrary to what is proposed in this paper, the prediction in JPEG-XR is performed between the coefficients, after the transform. Here we present a prediction scheme more similar to the one performed in H.264/AVC, in which, prior to the transform, the current pixels are predicted based on the previously encoded and reconstructed pixels.

The difficulty of our approach comes from the necessity to use for the prediction of the current block pixels in the causal neighborhood that have not been completely processed by the overlapping transform. We show how to overcome this difficulty, and thus the system presented here outperforms the traditional one, using DCT, in all tested images.

The remaining of this paper is organized as follows. In Sec. 2 we remind the state-of-the-art intra prediction in H.264/AVC and some basic facts about lapped transforms that will be used latter in the paper. Sec. 3 describes the proposed method for replacing block transforms by lapped ones and in Sec. 4 are provided some experimental results. We conclude in Sec. 5, giving also directions for further extensions, and application to video coding.

2. BACKGROUND

2.1. Lapped Transforms
In a general way, a lapped transform is any transform in which the filters are longer (L-tap) than the block being processed (M < L samples). Without any loss of generality, it is possible to express L as an integer multiple of M, L = NM, where N is the overlapping factor. Contrary to the block transforms, the matrix for the forward transform, P, is no longer square, but has M × L elements. For processing, as in the case of the block transforms, the signal x is divided into blocks. However, in this case, these new blocks are an extended version of the traditional ones containing M samples.
3. INTRA-FRAME PREDICTION AND LAPPED TRANSFORMS

Replacing the block transform with a lapped one after the intra-frame prediction is not an obvious task. The difficulty appears since the prediction is made in a closed loop, and it would be necessary to have the reconstruction of all the neighboring pixels in order to perform the prediction. However, it might be harder while using lapped transforms since several pixels which are not in the causal neighborhood are necessary for the reconstruction.

However, depending on the choice of the overlapping factor, it is possible to recover a part of the block. When using an overlapping factor of 2, for example, only the immediate neighbors are dependent and the filters overlap with only half of the pixels of the previous and next blocks. The other half can be reconstructed and be used to predict the next extended block \( \hat{v}_{nm} \), as can be seen in Fig. 2.a.

Since only half of the pixels on the left of the extended block are available, if one uses the traditional intra prediction, only four of the nine modes can be implemented. To improve this, and considering that many coders already divide the image in macroblocks of \( 16 \times 16 \) pixels, the order of encoding the block was changed, as illustrated in Fig. 2.b. With this new encoding order, half of the blocks (the blocks 1 and 3) in each macroblock would have all the necessary pixels for the prediction. On the other hand, for the blocks 2 and 4, only a few pixels would be available for the predictions, as seen in Fig. 2.c. In this case, it would only be possible to have the DC mode, an average of the available pixels.

Similarly to what is done in H.264/AVC, only the residual, \( R_{nm} \), difference between the prediction, \( \text{Pred}_{nm} \), and the actual extended block, \( v_{nm} \), will be transformed, quantized and then encoded. The process is detailed in the following equations. The direct 2D transform is carried out as

\[
Y_{nm} = P \cdot v_{nm} \tag{2}
\]

and the inverse transform defined through the matrix \( Q \),

\[
\hat{v}_{m} = Q^T \cdot Y_{nm} \tag{3}
\]

However, recovering the signal is not as direct as in the case of block transforms. The vectors \( v_{nm} \) and \( \hat{v}_{m} \) are different and the original signal is only obtained by accumulating the contribution of all blocks. Since any practical signal has finite length, a special border treatment [1], that will not be detailed here, has to be carried out for perfect reconstruction.

2.2. Intra-frame prediction

One of the techniques introduced in the H.264/AVC standard is the intra-frame prediction, which allows, while using a block transform, to take advantage of the redundancy of neighboring blocks. This is performed by using the pixels from previously encoded and reconstructed blocks to predict the current block. The prediction is then subtracted from the current block and only the residual (difference between the block and its prediction) and the prediction mode are transmitted to the decoder.

For the \( 4 \times 4 \) and \( 8 \times 8 \) blocks, there are nine intra-prediction modes in the H.264/AVC standard. Eight of them try to exploit possible directionality in the image. In the remaining mode (DC mode), all the block is predicted as the average of the neighboring pixels. In the case of \( 16 \times 16 \) blocks, only four modes are available. Three of them are similar to the modes availables for the other blocks DC, horizontal and vertical. The last one, called planar mode, consists in a smoothly-varying plane fitted to the upper and left pixels.

\[
SAD = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} |x_{nm}(i,j) - \text{Pred}_{nm}(i,j)| \tag{7}
\]
Fig. 2. Pixels available for the prediction (a) for the blocks are encoded in raster scanning order (b) for the blocks 1 and 3 in the new scanning order (c) for the blocks 2 and 4 in the new scanning order.

It is possible, while using lapped transforms, to have the same approach for the extended block, in which case all the pixels would receive the same significance. That would not be appropriate since they have different importance on reconstruction and on the rate. Alternatively, in order to improve that, we propose to weight differently each pixel of the extended $nm$-th residual block as follows:

$$W S A D = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} W_{ij} \left| v_{nm} (i,j) - \text{Pred}_{nm} (i,j) \right|,$$

where $W$ is a weighting matrix. In the tests presented later on, $W$ was given by

$$W = Q^T \cdot P \cdot \text{Ones} \cdot P^T \cdot Q,$$

where $\text{Ones}$ is an $L \times L$ matrix containing only ones. The matrix $W$ defined as above would correspond to the recovered extended block when all the pixels are equal to one. That may indicate the importance of each pixel in the reconstruction of the signal. Note that with this matrix, the pixels in the center of the block play a greater role in the reconstruction of the signal and typically the energy of the elements inside the currently encoded block represents more than 90% of the energy of the whole extended block.

4. EXPERIMENTAL RESULTS

Even though only gray scale images were considered in this work, the ideas presented here can easily be extended to color images. For simplicity, the tests were carried out in a JPEG-like coder. All the intra prediction modes were implemented and integrated.

The encoding block had to be fixed to $8 \times 8$ and the extended block to $16 \times 16$. Therefore, in the tests presented here, for the blocks 1 and 3, only the four proposed modes for $16 \times 16$ blocks have been implemented. On the other hand, for the DCT, all the nine modes have been implemented. That difference might give an advantage to the traditional scheme, but despite that, the results show that it can be outperformed by the proposed one.

The chosen lapped transform was GLBT $8 \times 16$ [6] (shorter transforms would impose further constraints on the filter design without any advantage for the implementation of intra prediction).

The encoding of the prediction modes with lapped transforms was performed differently than proposed in H.264/AVC, in which the most probable mode is predicted based in the neighbors. Since the other blocks in which the prediction is performed are not all immediate neighbors, it was not observed any advantage in using this approach. For the blocks 1 and 3, 2 bits were assigned to encode the mode. For the remaining blocks, 2 and 4, there is no need of encoding because only one mode is available.

The rate-distortion curves for two of the tested images can be seen in Figs. 3 and 4 where the direct application of lapped transforms (LT) performs better than traditional intra prediction with DCT (Intra+DCT), but it is outperformed by the presented scheme containing intra prediction with lapped transforms (Intra+LT). In Tab. 1, the average bit rate reduction and PSNR (in dB) difference are presented in Bjontegaard metric [15].

Fig. 3. Rate-distortion curves for the 512 × 512 pixels image Barbara
Fig. 4. Rate-distortion curves for the first frame of the full-HD sequence Riverbed

Table 1. Results for the proposed prediction method: gain G1 and G2 (dB) and rate reduction % R1 and % R2 compared to the H264-like DCT intra-prediction and direct application of lapped transform, respectively.

<table>
<thead>
<tr>
<th>Image (resolution)</th>
<th>G1(dB)</th>
<th>% R1</th>
<th>G2(dB)</th>
<th>% R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbara (512 × 512)</td>
<td>0.79</td>
<td>-11.21</td>
<td>0.31</td>
<td>-3.98</td>
</tr>
<tr>
<td>Goldhill (512 × 512)</td>
<td>0.51</td>
<td>-8.16</td>
<td>0.32</td>
<td>-4.57</td>
</tr>
<tr>
<td>Lena (512 × 512)</td>
<td>0.40</td>
<td>-4.07</td>
<td>0.35</td>
<td>-3.90</td>
</tr>
<tr>
<td>Riverbed (1080p)</td>
<td>1.12</td>
<td>-14.64</td>
<td>0.90</td>
<td>-4.72</td>
</tr>
<tr>
<td>Sunflower (1080p)</td>
<td>1.95</td>
<td>-7.65</td>
<td>1.48</td>
<td>-7.02</td>
</tr>
</tbody>
</table>

5. CONCLUSION AND DISCUSSION

The results presented here show that the intra-frame prediction scheme used in H.264/AVC can be efficiently adapted to be used with lapped transforms.

Previous results have already shown that lapped transforms could outperform the intra-frame prediction with DCT in H.264/AVC. However, those results were more consistent for transforms with long filters such as GenLot 8 × 48, which raise even more difficulties for the use with intra-frame prediction. In the case of transforms with short filters, such as the GLBT 8 × 16, used in this work, the results were more sensible with higher resolution images, which was confirmed in the tests presented here, as can be seen in the average PSNR gain. The scheme presented here outperforms the scheme using an traditional intra-frame prediction in all tested images even though fewer modes have been implemented for the moment. The implementation of all the nine modes (used in 4 × 4 and 8 × 8) for 16 × 16 blocks might improve the results presented here for intra frame coding.

Even though an important part of the gains presented here may be attributed to the application of the lapped transforms, the inclusion of an intra prediction, proposed in this work, presented consistent additional improvement in the performance in all tested images.

The extension of the prediction scheme presented here to video coding, as well as the implementation in a real H.264/AVC coder, is ongoing and further performance improvement is expected. Since the reference frames can be completely reconstructed, the difficulties presented here for the intra-frame prediction are not present in the inter-prediction.

6. REFERENCES