ATOMIC DECOMPOSITION DEDICATED TO AVC AND SPATIAL SVC PREDICTION

Aurélie Martin[1,2]*, Jean-Jacques Fuchs*[1]*, Christine Guillemot*[1] *and Dominique Thoreau*[2]

[1] IRISA - Université de Rennes 1 - Campus de Beaulieu - 35042 Rennes Cedex - France
[2] THOMSON Corporate research - 1 avenue Belle Fontaine BP 19 - 35510 Cesson-Sévigné Cedex - France

ABSTRACT

In this work, we propose the use of sparse signal representation techniques to solve the problem of closed-loop spatial image prediction. The reconstruction of signal in the block to predict is based on basis functions selected with the Matching Pursuit (MP) iterative algorithm, to best match a causal neighborhood. We evaluate this new method in terms of PSNR and bitrate in a H264/AVC encoder. Experimental results indicate an improvement of rate-distortion performance. In this paper, we also present results concerning the use of this technique for intra-inter layer prediction refinement, in a scalable video coding (SVC) like scheme.

***Index Terms*—** intra-prediction, atomic decomposition, extrapolation

1. INTRODUCTION

In H264/AVC, intra prediction is performed to decorrelate neighbor blocks. There are three intra prediction types called intra-16x16, intra-4x4 and intra-8x8 [1]. The prediction is based on the knowledge of the pixel row and column adjacent to the current block. Several directional modes are specified : four directions for intra-16x16 and eight directions for intra-4x4 and intra-8x8. The extrapolation is done by simply "propagating" the pixel values along one of the four or eight directions. Additionally to these geometrical modes, the DC prediction is available : this mode consists in predicting the current block from the mean of neighboring prior encoded samples. H264/AVC intra coding is very efficient to reconstruct uniform regions or directional structures, especially when one direction of intra modes best fits to the contours. However it is not possible to predict more complex textures. Alternative intra prediction methods based on block or template matching are suggested in [2] and [3] respectively.

To address the problem of signal prediction in highly textured areas, methods based on sparse signal approximations are considered here. The goal of sparse approximation techniques is to look for a linear expansion approximating the analyzed signal in terms of functions chosen from a large and redundant set (i.e. dictionary). The MP algorithm is a possible technique to compute adaptive signal representations by iterative selection of so-called *atoms* from the dictionary [4]. The MP algorithm has been later improved to give at each iteration the linear span of atoms which would give the best signal approximation in the sense of minimizing the residue of the new approximation. This improved algorithm is known as Optimized Orthogonal Matching Pursuit (OOMP) [5]. The MP algorithm has been applied to low rate video coding in [6]. Motion residual images are decomposed into a weighted summation of elements from a large dictionary of 2-D Gabor structures. Used with a time-frequency dictionary of Gabor functions MP provides a high-resolution adaptive parametrization of signal's structures. MP has also been applied to signal extension using cosines and wavelet basis functions [7]. Here, we consider the problem of closed-loop spatial image prediction or extrapolation. It can be seen as a problem of signal extension from noisy data taken from a causal neighborhood. The MP sparse representation algorithm is considered. We also present a way to improve upsampled images in the context of SVC coding thanks to atomic decompositions.

The remainder of the article is organized as follows. The MP algorithm is first recalled in section 2. The adaptation of this algorithm to the prediction problem is presented in section 3. The application consisting in a refinement of inter-intra layer in the SVC scheme is explained in section 4. In section 5.1, the MP prediction is compared against AVC prediction and in section 5.2, SVC refined prediction results are presented.

2. MATCHING PURSUIT ALGORITHM (M.P.)

Let Y be a vector of dimension N and A a matrix of dimension $N \times M$ with $M \gg N$. The columns a_k of A can be seen as basis functions or atoms of a dictionary that will be used to represent the vector Y. Note that there is an infinite number of ways to choose the M dimensional vector X such that $Y = AX$. The aim of sparse representations is to search among all these solutions of $Y = AX$ those that are sparse, i.e. those for which the vector X has only a small number of nonzero components. Indeed one quite generally does not seek an exact reconstruction but rather seeks a sparse representation that satisfies

$$\|Y - AX\|_2^2 \leq \rho$$

where ρ characterizes an admissible reconstruction error. Since searching for the sparsest representation satisfying this constraint is NP-hard and hence computationally intractable, one seeks approximate solutions. The MP algorithm offers a sub-optimal solution to this problem via an iterative algorithm. It generates a sequence of M dimensional vectors X_k having an increasing number of non zero components in the following way. At the first iteration $X_0 = 0$ and an initial residual vector $R_0 = Y - AX_0 = Y$ is computed. At iteration k, the algorithm selects the basis function a_{j_k} having the highest correlation with the current residual vector $R_{k-1} = Y - AX_{k-1}$, that is, such that

$$j_k = \arg\max_j \frac{\left(a_j^T R_{k-1}\right)^2}{a_j^T a_j}.$$

The weight x_{j_k} of this new atom is then chosen so as to minimize the energy of the new residual vector, which becomes thus equal to

$$R_k = R_{k-1} - \frac{a_j^T R_{k-1}}{a_j^T a_j} a_{j_k}.$$

The new optimal weight is introduced into X_{k-1} to yield X_k Note that the same atom may be chosen several times by MP. In this case, the value of the coefficient is added to the previous one. The algorithm proceeds until the stopping criterion

$$\|Y - AX_k\|^2 \leq \rho \tag{1}$$

is satisfied, where ρ is a tolerance parameter which controls the sparseness of the representation.

3. PREDICTION BASED ON MP

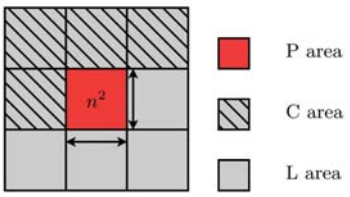

Fig. 1. C is the causal area, P is the current block to be predicted and L is the whole area surrounding P

In Fig. 1, we define the block P of $n \times n$ pixels to be predicted using its causal neighborhood C of size $4n^2$. With the entire region L containing 9 blocks and hence of size $3n$x$3n$ pixels, we associate the Discrete Fourier and/or Cosine basis functions expressed respectively as

$$g_{p,q}(m,n) = e^{2i\pi(\frac{mp}{M}+\frac{nq}{N})} \tag{2}$$

and

$$g_{p,q}(m,n) = cos\left(\frac{(2m+1)p\pi}{2M}\right) cos\left(\frac{(2n+1)q\pi}{2N}\right). \tag{3}$$

With these atoms we build the matrix A. In the experiments reported in section 4, this matrix is composed of $9n^2$ atoms (DCT or DFT) or $18n^2$ atoms (DCT and DFT), however it can be extended to include other basis functions as for instance Gabor or wavelets. We denote Y the $9n^2$ dimensional vector formed with the pixel values of the area L and X the vector containing the coefficients of the representation of Y in terms of the basis functions : $Y = AX$. The matrix A is modified by masking its rows corresponding to the pixels not in the known area C. We thus obtain a compacted matrix A_c whose size is $4n^2$x$9n^2$ if only the DCT basis is considered. The corresponding components in Y are deleted similarly to get the vector Y_c of $4n^2$ pixels. The MP algorithm is then applied to A_c and Y_c. For later use, we define similarly A_p and Y_p of size n^2x$9n^2$ and n^2x1 associated with the area P to be predicted.

Remember that the aim of MP algorithm is to get a sparse representation of Y_c. This means that as the complexity of the representation i.e. as the number k of non zero components in X, increases the reconstruction error

$$\|Y_c - A_c X_k\|^2 \tag{4}$$

decreases monotonically. Here, X_k denotes the representation proposed by the MP algorithm after k steps.

But since our purpose is to get a good prediction of the area P there is of course no reason that the better the representation of the area C, the better the associated prediction of the area P. We will therefore apply to MP a stopping criterion that tends to fulfil this goal, i.e., that tends to minimize the reconstruction error in P. We implement the algorithm so that it generates a sequence of representations X_k of increasing complexity and for each X_k we compute the prediction error energy $\|Y_p - A_p X_k\|^2$ and we should thus stop as soon as this prediction error which generically starts decreasing, increases. But since there is no reason that a more complex representation cannot indeed yield a smaller prediction error, we actually proceed differently and consider a two steps procedure.

First the MP algorithm are run until the pre-specified threshold on the reconstruction error in (4) is reached and the resulting X_k sequences are stored. The values of the thresholds are fixed such that the final representation has a quite large number of components, say K. In a second step one then selects the optimal representation as the one that gives the smallest error energy on the area P to be predicted:

$$k_{opt} = \min_{k \in [1,\ K]} \|Y_p - A_p X_k\|_2^2 \tag{5}$$

The optimal number of atoms k_{opt} is transmitted to the

decoder side in order to be able to compute the same prediction.

4. SPARSE REPRESENTATION FOR INTER-INTRA LAYER PREDICTION

4.1. Brief introduction to SVC

Scalable Video Coding scheme has been approved as an extension of H264/AVC standard since July, 2007 [8]. This new video codec produces bitstreams decodable at different bitrates. SVC provides a large degree of flexibility in terms of scalability :

◇ *Temporal scalability* : the video can be encoded at different temporal frequencies.

◇ *Spatial scalability* : spatial adaptability leads to multiple resolutions. HD-videos for instance, can be decoded into SD, CIF or QCIF format.

◇ *Quality scalability or SNR scalability* : the amount of information to reconstruct the signal can be chosen according to the fiability of the channel.

We will focus here on improving the spatial scalability performance.

4.2. Spatial scalability in SVC

SVC is a layered video codec which induces spatio-temporal and quality scalability. To improve the compression efficiency, the aim is to find the best prediction image to reduce the energy of the residual error signal. SVC standard bases its spatial prediction on dependencies between different resolutions. The main goal is to use as much as possible base layer information. Each current macroblock is predicted with up-sampled lower resolution signal. To improve coding efficiency, three upsampling process are combined : texture, residual and motion upsampling.

The upsampling process on luminance component, texture upsampling, is performed with one dimensional 4-taps FIR filters, horizontally and vertically, on intra predicted blocks. The chrominance components are upsampled with bilinear filters. The encoder also performs an intra prediction thanks to the directional modes of H264/AVC standard. The best predictor (e.g. lagrangian criterion) is then selected. If the current block corresponds to an inter-coded macroblock in the base layer, the enhancement layer macroblock is inter-coded. The inter-layer residual prediction can also be employed for inter or intra coded blocks to improve scalable coding.

4.3. Spatial SVC prediction with MP

The main goal of our approach is to take advantage of two predictions : intra and spatial inter-layer prediction. As showed in section 3, the information of neighbor pixels previously reconstructed is, in most cases, sufficient to recollect the unknown current block. When the signal to predict corresponds to new patterns or appearing edges, the reconstruction is quite impossible. The best chosen prediction is then the mean of the reconstructed causal area. In SVC, the current picture at a lower resolution, the basis layer, is available. The current block is not yet predicted so non-causal components of the upsampled basis layer can be used in addition to adjacent pixels in the current picture. These pixels represent a major source of information, especially when the signal is not predictable with only neighbor pixels. We modify the algorithm to take upsampled basis layer pixels into account, as depicted on Fig. 2. Pixels surrounding the current block are still considered and we also insert pixels of the upsampled basis layer in the input vector Y. As supplementary information about the current block to predict is available, the algorithm is expected to have better performances. Actually, the stop criterion (1) of MP algorithm is much more efficient because the constraints are not *only* on neighbor pixels any more. Each selected atom is also adapted to the part of signal extracted of the upsampled image.

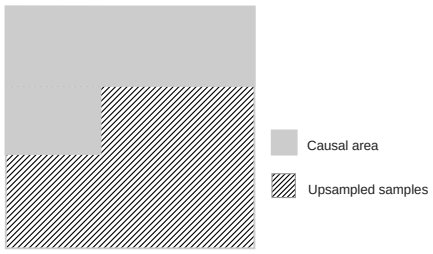

Fig. 2. Prediction "refinement"

5. SIMULATION RESULTS

5.1. Results about MP prediction

We consider the spatial prediction of blocks of 4×4, 8×8, and 16×16 pixels ($n = 4$, 8 or 16). The Cosine functions have been used to construct the redundant dictionary A. The threshold is set to a value that yields a final representation having K, a quite large number, of non zero components. Then the vector X related to the optimal representation is selected, see (5). In all our simulations ρ is set to 1 in (1). The MP based prediction was integrated in JM 11.0 KTA 1.2 (Key Technical Area) software without any change of the encoder syntax. The proposed prediction mode substitutes for one AVC mode for each type of prediction (intra-4x4, 8x8 or/and 16x16). The selected AVC mode corresponds to the less chosen mode. Results concerning the following tests are presented : the MP based prediction subsitutes one AVC mode when the three prediction types were combined, or when only intra-4x4 and 8x8 were available or just intra-4x4. Note that an additional flag was inserted to turn intra-16x16 prediction off. For instance, mode 6 (horizontal down) for

intra-8x8 and 4x4 was replaced by our sparse representation, for the Barbara picture. Simulations were performed on a large range of quantization levels to evaluate the Bjontegaard (BJ) average PSNR improvement of luminance components and bit rate savings. Table 1 presents the results for MP prediction according to the three types of intra-prediction. The higher rate savings are obtained when intra-4x4 and intra-8x8 prediction are combined. Note that the cost for encoding the number of coefficients, is not taken into account.

	4x4		4x4, 8x8		4x4, 8x8, 16x16	
QP	psnr	rate (kb)	psnr	rate (kb)	psnr	rate (kb)
15	46.69	27 535	46.51	26 677	46.52	26 680
25	38.66	11 541	38.69	11 303	38.69	11 285
35	31.43	4 562	31.80	4 260	31.75	4 196
45	24.65	1 768	25.97	1 464	25.84	1 335
	4x4		4x4, 8x8		4x4, 8x8, 16x16	
BJ	+ 0.64 dB -7.55 %		+ 0.63 dB -8.25 %		+ 0.57 dB -7.70 %	

Table 1. Bjontegaard results for MP implemented in KTA-software; QP is the quantization parameter

5.2. Results for the SVC spatial approach

These results present a spatial inter-layer MP based prediction, see 4.3. The algorithm runs both with the knowledge of surrounding reconstructed pixels and upsampled components from a lower resolution. To generate the quantized basis layer, we first create a lower resolution picture thanks to a gaussian pyramid scheme. The downsampling filter used is the following AVC filter: [26 19 5 -3 -4 0 2]. Then this image is encoded in the KTA software and upsampled with AVC upsampling filter [20 -5 1].We chose to set the quantization parameter to the same value for the basis layer and the current image. Besides, the encoder runs with the MP prediction but the difference is that non causal area surrounding the current block is filled with pixels of the upsampled image of the basis layer. In order to compare our method against the usual inter-intra layer SVC spatial prediction, we establish a reference, independently of the previous test. One substitutes the prediction block corresponding to the upsampled image, to the statistically less used intra AVC mode (mode 6). We also chose to introduce the prediction from the upsampled basis layer instead of the second statistically less chosen AVC mode (mode 8). Fig. 3 shows the performance of our prediction put in competition with the upsampled basis layer (UBL). The AVC upsamlping filter well suits to recover contours between two different textured areas. However, at a high quantization level, the texture is lost. Therefore the MP algorithm offers an interesting alternative to recollect the signal.

6. CONCLUSIONS

This new approach of intra prediction offers interesting perspectives compared to directional modes of H264/AVC. For

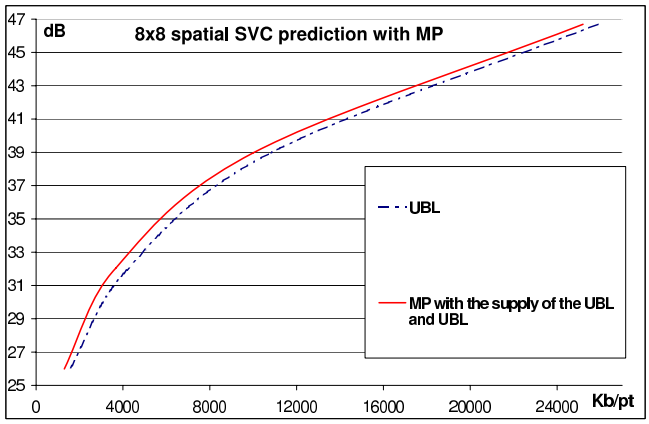

Fig. 3. MP+UBL performance - Associated Bjontegaard results : + 0.75 dB , - 9.85 %

complex textures, the MP algorithm turns out to be an interesting alternative for Intra prediction and also for inter-intra layer prediction. Concerning this second application, the knowledge of non causal components allows the MP algorithm to extrapolate local textures more precisely. Simulation results show a bitrate improvement of 8 % in AVC and nearly 10 % in SVC. Further works will be focused on the adaptation of the stop criterion, highly dependant on the dynamic of the surrounding signal.

7. REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC" *IEEE Trans., Vol 13,7, 560 - 576*, July 2003

[2] J. Yang, B. Yin, Y. Sun and N. Zhang, "A block-matching based intra frame prediction H.264/AVC" *ICME*,2006.

[3] T. K. Tan, C. S. Boon and Y. Suzuki, "Intra prediction by template matching" *ICIP*,2006.

[4] S. Mallat and Z. Zhang, "Matching Pursuits with time frequency dictionaries" *IEEE Sig. Processing*,vol. 41, 12, dec 1993.

[5] G.M. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations", *Conts. Approx.*, Vol 13, 57-98, 1997.

[6] R.Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuit video coder", *IEEE Circuits and systems for video technology*, vol. 7, 1, feb. 1997.

[7] U.T. Desai, "DCT and Wavelet based representations of arbitraily shaped image segments", *proc. IEEE Intl. Conference on Image Processing*, 1995.

[8] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding extension of the H264/AVC standard", *IEEE Special issue on SVC*, Vol. 17, No 9, pp 1103-1120, sept 2007.