15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007, copyright by EURASIP

SPARSE REPRESENTATION FOR IMAGE PREDICTION

Aurélie Martin, Jean-Jacques Fuchs, Christine Guillemot and Dominique Thoreau

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

ABSTRACT

This paper addresses the problem of closed-loop spatial image prediction based on sparse signal representation techniques. The basis functions which best approximate a causal neighborhood are used to extrapolate the signal in the region to predict. Two iterative algorithms for sparse signal representation are considered: the Matching Pursuit algorithm and the Global Matched Filter. The predicted signal PSNR achieved with these two methods are compared against those obtained with the directional predictive modes of H.264/AVC.

1. INTRODUCTION

Closed-loop spatial prediction has been widely used for image compression in transform (H.261/H.263, MPEG-1/2/4) or spatial (H.264) domains. In H.264, there are two intra prediction types called Intra-16x16 and Intra-4x4 respectively [1]. The Intra-16x16 type supports four intra prediction modes while the Intra-4x4/8x8 type supports nine modes. Each 4x4 block is predicted from prior encoded samples from spatially neighboring blocks. In addition to the so-called "DC" mode which consists in predicting the entire 4x4 block from the mean of neighboring pixels, eight directional prediction modes are specified. The prediction is done by simply "propagating" the pixel values along the specified direction. This approach is suitable in presence of contours, the directional mode chosen corresponds to the orientation of the contour. However, it fails in more complex textured areas. 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 (dictionnary). The Matching Pursuit (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]. An alternative optimal sparse representation called Global Matched Filter (GMF) has been described in [11]. The advantage of GMF, compared to MP, is that the best *atoms* are simultaneously selected instead of choosing them one by one.

The MP algorithm has been applied to low rate video coding in [7]. 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 [8].

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. Both the MP and the GMF sparse representation algorthims are considered. The sparse signal approximation is run with a set of *masked* basis functions, the masked samples corresponding to the location of the pixels to be predicted. However, the stopping criterion (which is the energy of the residue) is computed on the region to predict. To compute it on the causal neighborhood would lead to a residue of small energy, however, this residue might take potentially large values in the region to be predicted. The number of *atoms* selected in order to minimize the energy of the residue on the region to predict is transmitted. The decoder runs the algorithm with the *masked* basis functions and taking the previously decoded neighborhood as the known support. The number of atoms selected by the encoder is used by the decoder as a stopping criterion.

The remainder of the article is organized as follows. The MP and the GMF algorithms are first recalled in sections 2.1 and 2.2 The adaptation of these algorithms to the prediction problem is presented in section 3. The approaches are compared against the H.264 prediction modes in section 4.

2. SPARSE REPRESENTATIONS

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 recconstruction 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.

2.1 Matching Pursuit algorithm (M.P.)

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

$$j_k = \arg\max_j \frac{|a_j^T R_{k-1}|}{a_j^T a_j}.$$

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

$$\|Y - AX_k\|^2 \leq \rho \qquad (1)$$

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

2.2 Global Matched Filter (G.M.F.)

The G.M.F introduced in [6] is an interesting alternative to the MP algorithm. Keeping the same notation as in section 2.1, the GMF algorithm yields the sparse representation that minimizes the criterion

$$\min \frac{1}{2} \|Y - AX\|_2^2 + h\|X\|_1 \qquad \text{with} \quad h > 0 \qquad (2)$$

where $\|X\|_1 = \sum |x_j|$ and $h > 0$ is a threshold which controls the sparseness of the representation. Indeed (2) can be rewritten as

$$\min \|AX\|_2^2 \qquad \text{subject to} \quad \|A^T(Y - AX)\|_\infty \leq h,$$

where the constraint can be given the following physical interpretation. At a point, say $\hat{X}$ satisfying it,

- $\diamond$ $Y - A\hat{X}$ is the residual vector which can be seen as the unexplained part Y,
- $\diamond$ $A^T(Y - A\hat{X})$ is a vector containing all the correlations of the atoms with the residual vector,
- $\diamond$ the constraint guarantees that, at an admissible point $\hat{X}$, no component in this vector exceeds h.
- $\diamond$ one finally seeks the representation satisfying this constraint that has minimal energy.

At the optimum all the used atoms have indeed their correlation equal to h. As opposed to MP which is an ad hoc procedure, GMF is optimal. The advantage of GMF, compared to MP, is that for one value of h, the best elements of A are simultaneously selected, instead of choosing the atoms one by one.

The price to pay for this difference is a higher computational burden, although there are now quite efficient ways to implement GMF. The criterion (2) has appeared somehow simultaneously in different communities and is sometimes known as Basis-Pursuit denoising [9]. An efficient manner to implement the GMF, that we will be using in the sequel, has been developed in the statistics community [10] (see also [11]). Although one wants to solve (2) for a fixed h, the algorithm works iteratively in the number of components (just as MP) and starts with $h = \left\|A^T Y\right\|_\infty$ for which a first non zero component appears in the component of the optimal X. The value of h is then decreased and the next value of h for which a second component becomes non zero is found; The algorithm proceeds in this way until the desired value of h falls within the current interval in h. One searches (adjacent) intervals of h within which - the number of nonzero components in the optimum X remains constant and - an explicit expression of the optimum is known. Using this quite efficient algorithm has indeed a further advantage : it allows to build a sequence, say X_k of optimal representations with increasing complexity. The basic MP algorithm proceeds in the same way but at step k, the X_k it generates has no real optimality property.

Remarks: In the sparse representation context, it is important to note that the larger M, the number of components (atoms) in the redundant basis (dictionary), the smaller the number of components required in a potential "good" representation, but also the higher the computational complexity. Hence the importance to choose a good dictionary or better to adapt it to the signals to be represented. In the sequel we will essentially use the Discrete Cosine Transform basis and the Discrete Fourier Transform basis in the real pixel-domain, i.e., the vector Y is filled with pixels of the available blocks of the image under investigation.

3. PREDICTION BASED ON MP AND GMF

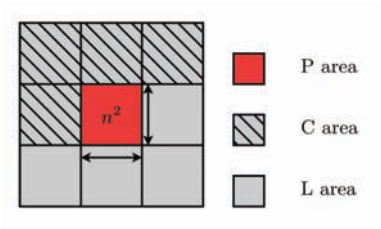

Figure 1: C is the causal area, P is the current block to be predicted and L is the whole area surrounding P

In Figure 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

$$g_{p,q}(m,n) = e^{2i\pi(\frac{mp}{M} + \frac{nq}{N})} \qquad (3)$$

and

$$g_{p,q}(m,n) = cos\left(\frac{(2m+1)p\pi}{2M}\right) cos\left(\frac{(2n+1)q\pi}{2N}\right). \qquad (4)$$

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 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 and GMF algorithms are then applied to A_c and Y_c.

Remember that the aim of both algorithms 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

$$\|Y_c - A_c X_k\|^2$$

decreases monotonically. Here, X_k denotes both the representation proposed by the MP algorithm after k steps and the optimum of the GMF criterion (2) for an h which would be chosen such that there are k non zero components in this optimum.

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 both MP and GMF a stopping criterion that tends to fulfil this goal, i.e., that tends to minimize the reconstruction error in P. We implement both algorithms so that they generate 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 and the GMF algorithms are run until the pre-specified threshold 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 $k_{\max}$. 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:

$$k_{opt} = \min_{k \in [1,\ k_{\max}]} \|Y_p - A_p X_k\|_2^2 \qquad (5)$$

The optimal number of atoms k_{opt} is transmmitted to the decoder side and this allows him to compute the same prediction.

4. SIMULATION RESULTS

We consider the spatial prediction of blocks of 8×8 pixels ($n = 8$). Both discrete real Fourier and Cosine functions have been used to construct the redundant dictionnary A. The real Fourier basis functions are defined as:

$$f(n) = \frac{1}{N}\{F(0) + \sum_{p=0}^{N/2-1} (F(p)e^{2i\pi\frac{pn}{N}} + F^*(p)e^{-2i\pi\frac{pn}{N}}) \qquad (6)$$
$$+(-1)^n F(\frac{N}{2})\}.$$

where f is the 1D real function of support of length N computed by taking the inverse Fourier transform of the vector including the known and the unknown samples (F denotes the Fourier transform). The corresponding 2D basis functions are obtained with the Kronecker product.

Lena	PSNR	Nb of components
MP	29.66	3
GMF	31.54	3
Barbara	PSNR	Nb of components
MP	30.88	19
GMF	31.12	15
House	PSNR	Nb of components
MP	31.53	5
GMF	33.10	20

Table 1: Results with the real DFT

Each threshold is set to a value that yields a final representation having $k_{\max}$, 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 we have taken $\rho = 8$ in (1) and $h = 8$ in (2). Tables 1 and 2 first compare the respective performance of MP and GMF in terms of the PSNR of the predicted signal and of the number of basis functions selected, considering real Discrete Fourier basis functions and Discrete Cosine functions. It can be observed that there is a performance gap between 1dB and 3.8 dB depending on the image. Table 3 shows results with an extended dictionary filled with both the Fourier basis and the cosine basis functions. Overall, the combination of two basis leads to higher performances since the redundancy of the dictionary increases.

Lena	PSNR	Nb of components
MP	29.41	2
GMF	31.65	12
Barbara	PSNR	Nb of components
MP	27.00	12
GMF	29.85	11
House	PSNR	Nb of components
MP	32.28	4
GMF	34.28	40

Table 2: Results with the DCT

Lena	PSNR	Nb of components
MP	29.41	2
GMF	31.75	41
Barbara	PSNR	Nb of components
MP	34.06	17
GMF	34.29	11
House	PSNR	Nb of components
MP	33.23	3
GMF	33.74	18

Table 3: Results with a larger dictionary that includes the DCT and the DFT

The two algorithms are then compared against the 9 directional prediction modes of H.264. They are then used as additional modes to the 9 directional prediction modes of H.264. The prediction mode giving the lowest sum of squared error ($SSE = \sum_i (f_i - p_i)^2$, where f_i is one pixel of the reconstructed image and p_i one of the predicted image) is selected. Tables 5 and 6 give the PSNR of the predicted image obtained with the different modes and the percentage of selection of the different modes. The last line of Tables 4, 5 and 6 gives the mean PSNR value of the image predicted with the modes that minimize the SSE. It can be observed that mode 9 (corresponding to MP in Table 5 and to GMF in Table 6) is the most selected and leads to the highest PSNR. Fig 2 shows the 8x8 blocks, colored in blue, that have been predicted using MP and GMF. The 9 directional modes of H.264 remain selected in smooth areas whereas MP and GMF are selected in highly textured regions. As stated above, the number of coefficients (not the coefficients themselves) needs to be transmitted. However, this overhead remains low. To see the improvements of GMF compared to the 9 directional modes of H264, refer to Tables 4 and 6. A gain of 1.2 dB is obtained by including GMF as an extra prediction mode.

Fig 3 and Fig 4 show that directional modes still perform a better reconstruction on smooth areas and along simple directions. But GMF is much better on textured area.

Fig 3 (right) shows the predicted image obtained with the best modes among 10 (last line in Table 6)

Fig 5 and Fig 6 show the positions of non-zero coefficients for the MP and the GMF algorithms. It can be seen that, overall, the GMF algorithm leads to a sparsiest representation of the signal.

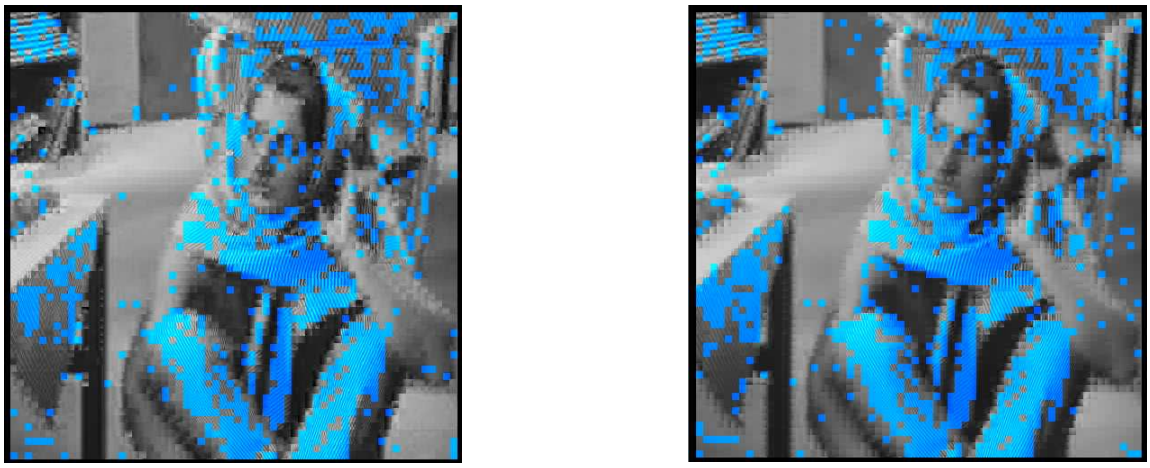

Figure 2: Image predicted with MP (left), image predicted with GMF (right). Blocks in blue are those selected withe the SSE criterion, compared to the 9 modes of H264.

Figure 3: Image predicted with the 9 directional modes of H264/AVC (left), image predicted with AVC modes and GMF (right)

Figure 4: Detail from the image predicted with the 9 directional modes of H264/AVC (left), image predicted with AVC modes and GMF (right)

modes	percentage	psnr
mode 0	19.667	20.280
mode 1	8.897	18.436
mode 2	12.435	20.309
mode 3	7.284	18.889
mode 4	7.778	19.035
mode 5	14.178	19.686
mode 6	5.099	18.598
mode 7	16.415	20.067
mode 8	8.247	18.422
best modes		23.375

Table 4: Results with the 9 modes of H264

modes	percentage	psnr
mode 0	16.363	20.280
mode 1	7.466	18.436
mode 2	7.440	20.309
mode 3	5.411	18.889
mode 4	6.322	19.035
mode 5	10.250	19.686
mode 6	4.683	18.598
mode 7	11.837	20.067
mode 8	7.206	18.422
mode 9	**23.023**	**21.997**
best modes		24.441

Table 5: Results with MP (mode 9) and the 9 modes of H264.

modes	percentage	psnr
mode 0	15.973	20.280
mode 1	7.102	18.436
mode 2	6.842	20.309
mode 3	5.073	18.889
mode 4	6.035	19.035
mode 5	9.469	19.686
mode 6	4.370	18.598
mode 7	10.900	20.067
mode 8	7.154	18.422
mode 9	**27.081**	**22.202**
best modes		24.568

Table 6: Results with GMF (mode 9) and the 9 modes of H264

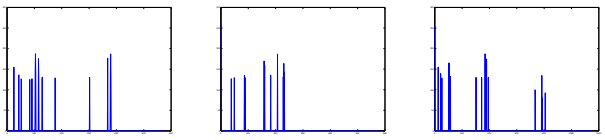

Figure 5: Coefficients selected by the MP algorithm with the DFT (left), the DCT (middle) and both of the transforms : DCT and DFT (right)

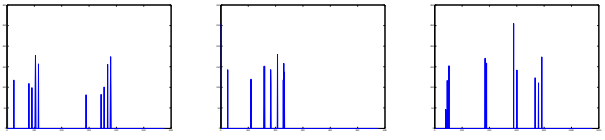

Figure 6: Coefficients selected by the GMF algorithm with the DFT (left), the DCT (middle) and both of the transforms : DCT and DFT (right)

5. CONCLUSIONS

This new approach of intra prediction offers interesting perspectives compared to directional modes of H264/AVC. The nine modes of the standard still performed the best prediction on smooth areas. But for complex textures, both MP and GMF algorithms turn out to be an interesting alternative. The GMF showed optimized results especially where textures have high variations, that are generally difficult to reconstruct. Therefore, MP and GMF can avantageously substitute for one directional mode. Future effort will be dedicated to an assessment of the rate-distortion performance of the prediction techniques in the H.264 video codec.

REFERENCES

[1] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC" *Circuits and Systems for Video Technology, IEEE Transactions, 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] J.J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays" *IEEE Sig. Processing*,vol. 49, 4, april 2001.

[7] 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.

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

[9] S. Chen, D. Donoho and M. Saunders, "Atomic Decomposition by Basis Pursuit" *SIAM J. on Scientific Comput.*, 20, 1, 33-61, 1999.

[10] B. Efron, T. Hastie, I. Johnstone and R.Tibshirani, "Least angle regression," *Annals of Statistics*, 32, pp. 407–499, Apr. 2004.

[11] S. Maria and J.J. Fuchs, "Application of the global matched filter to stap data, an efficient algorithmic approach" *ICASSP*,2006.