NEW DICTIONARY AND FAST ATOM SEARCHING METHOD FOR MATCHING PURSUIT REPRESENTATION OF DISPLACED FRAME DIFFERENCE

Fulvio Moschetti[1], Lorenzo Granai[1], Pierre Vandergheynst[1], Pascal Frossard[2]

[1] ITS Institute of Signal Processing EPFL, CH 1015 Lausanne Switzerland
[2] IBM TJ Watson Research Center PO Box 219, Hawthorne, NY 10598, USA

ABSTRACT

Matching Pursuit decomposes a signal into a linear expansion of functions selected from a redundant dictionary, isolating the signal structures that are coherent with respect to a given dictionary. In this paper we focus on the Matching Pursuit representation of the displaced frame difference (dfd). In particular, we introduce a new dictionary for Matching Pursuit that efficiently exploits the signal structures of the dfd. We also propose a fast strategy to find the atoms exploiting the max of the absolute value of the error in the motion predicted image and the convergence of the MSE with the rotation of the atoms.

Results show that the fast strategy is quite robust when compared to exhaustive search techniques and it improves the results of a suboptimal search strategy based on a genetic algorithm.

1. INTRODUCTION

High compression ratios in video coding are achieved by adopting hybrid systems that combine two stages. In the first stage motion estimation and compensation predict each frame from the neighboring frames. At the second stage the prediction error is coded. Current video compression standards use block based DCT to code the residual error. In [1, 2] authors have shown that improved coding efficiency can be achieved by replacing the DCT with an overcomplete transform. Non orthogonal transforms represent indeed a valid alternative to orthogonal transforms like DCT or wavelet based scheme especially at low bit-rates, where most of the signal energy can be captured by only few elements. Matching Pursuit (MP) algorithms iteratively decompose a signal in its most important features using a set of atoms chosen among a redundant dictionary of basis functions. Particular attention has to be dedicated to the dictionary design since it impacts the coding performance.

In this paper we present a new dictionary that efficiently captures the contour and edges of dfd images. It improves the performance of a previously introduced dictionary [3], based on oriented and anisotropically refined atoms, whose quality performance already overwhelmed the commonly used two dimensional separable Gabor function.

The main limitation in the adoption of a redundant dictionary remains the encoding complexity. For this reason we propose a fast and efficient method for atom selection. This method has been compared to another sub-optimal approach based on a genetic algorithm and to a full search based approach.

2. MATCHING PURSUIT

A detailed explanation of the theory of the Matching Pursuit can be found in Mallat e Zhang [4]. Here we just recall the basics of the iterative process used for the selection of the waveforms that represent the signal structures.

Let $D = \{g_\gamma\}_{\gamma \in \Gamma}$ be a dictionary of unitary norm vectors g_γ called atoms and Γ represent the set of possible indexes. The function f is first decomposed as follows:

$$f = \langle g_{\gamma_0}, f \rangle g_{\gamma_0} + Rf, \qquad (1)$$

where Rf is the residual component after having approximated f in the direction of g_{γ_0}.

Since Rf and g_{γ_0} are orthogonal, it follows that

$$\|f\|^2 = \left| \langle g_{\gamma_0}, f \rangle \right|^2 + \|Rf\|^2. \qquad (2)$$

To minimize $\|Rf\|$, we must choose g_{γ_0} such that the $\left| \langle g_{\gamma_0}, f \rangle \right|$ projection is maximal. Applying iteratively such a procedure, after N iterations we obtain:

$$f = \sum_{n=0}^{N-1} \langle g_{\gamma_n}, R^n f \rangle g_{\gamma_n} + R^N f \qquad (3)$$

where $R^n f$ is the residual at the nth step and $R^0 f = f$. As in (2) we can write

$$\|f\|^2 = \sum_{n=0}^{N-1} \left| \langle g_{\gamma_n}, R^n f \rangle \right|^2 + \|R^N f\|^2. \qquad (4)$$

This equation expresses the energy conservation.

The convergence of MP depends on both the dictionary and the (sub)optimal search strategy. In [4] it has been shown that there are two real numbers $\alpha, \beta \in \,]0,1]$ such that for all $m \geq 0$ the following relation is valid:

$$\left\| R^{m+1} f \right\| \leq \left(1 - \alpha^2 \beta^2 \right)^{1/2} \cdot \left\| R^m f \right\|. \qquad (5)$$

α is an optimality factor related to the strategy adopted to determine the best atom in the dictionary, while β strictly depends on the dictionary representing its ability to capture the feature of the input function f [5].

This dictionary is built acting on a generating function of unit L^2 norm by means of a family of unitary operators $U\gamma$:

$$D = \{ U_\gamma, \gamma \in \Gamma \}, \qquad (6)$$

for a given set of indexes Γ. Basically this set contains three types of operations: translation $\vec{d}$, rotation θ and anisotropic scaling (c_1, c_2).

A possible action of $U\gamma$ on the generating atom g is thus given by:

$$U_\gamma g = U(\vec{d}, \theta) A(c_1, c_2) g \qquad (7)$$

where U is an operator that acts on translation and rotation, while A is an anisotropic dilation operator.

The system we implemented for the Matching Pursuit representation takes as input the difference between the original frame and the motion compensated one. Motion estimation-compensation is realized as full search block matching. Matching Pursuit decomposition is then applied on blocks SxS, with S=32. The decomposition is halted when MSE reaches a threshold value of 65, or when the number of atoms exceed 7 per block.

3. NEW DICTIONARY

In our case we are targeting a particular set of images, which are the by product of the motion compensation. They have a particular signal structure, characterized by edges. To capture these features the proposed dictionary is composed of anisotropically refined atoms. Anisotropy increases the redundancy in the dictionary because of the introduction of an extra parameter to code, but as it has been shown in [3] this produces an overall increase in efficiency. In our experiments [6] we chose as generating function, the combination of a gaussian and a triangular function, that is

$$g_0 (x, y) = w_0 \cdot e^{-\left(x^2 + y^2 \right)}, \text{ where} \qquad (8)$$

$$w_0 = \begin{cases} 2 - 4|x| & \text{for } 0 \leq |x| < 1 \\ -4 + 2|x| & \text{otherwise} \end{cases}.$$

We have compared this dictionary with the one introduced in [3], that uses the combination of a gaussian and its second derivative, as expressed by :

$$g_1 (x, y) = \left(4x^2 - 2 \right) \cdot e^{-\left(x^2 + y^2 \right)} \qquad (9)$$

A possible way to compare two dictionaries consists in considering the convergence speed of Matching Pursuit, that corresponds to its ability to extract the maximum signal energy in a few iterations. Namely, the decay rate of the residue represents the coding efficiency of the Matching Pursuit.

From (5) we can see that the error decay rate involves two parameters α and β. Using an exhaustive searching strategy of the parameters of the atoms, α becomes not influential anymore (since equal to 1), so we can try to estimate β, which depends solely on the dictionary construction. In this case we have:

$$\beta^2 \leq 1 - \frac{\left\| R^{m+1} f \right\|^2}{\left\| R^m f \right\|^2} \qquad (10).$$

Eq. (10) sets an upper limit for β which can be estimated measuring the values $T_m = \dfrac{\left\| R^{m+1} f \right\|^2}{\left\| R^m f \right\|^2}$ for each iteration m.

We can also notice that the efficiency of the dictionary might change at each iteration, as a result of the changing nature of the residue in the iterative Matching Pursuit operation. For three different sequences (Stefan, Akiyo and Coastguard) of five images, we have measured the values of T_m at each iteration for all the blocks into which the image has been divided. We have estimated in this way the value of β. Results show an improvement for the dictionary obtained using g_0 rather than g_1 of about 4%, with values of β respectively 0.50 and 0.48.

These results represent an upper bound in the decay of the residual images. In section 5 we will see that in the practical case the spread between the two dictionaries is more pronounced.

4. NEW FAST CONVERGING ALGORITHM

The direct application of the Matching Pursuit algorithm would require us to test each 2-D basis function at all possible integer-pixels locations in the image and compute all of the resulting inner products.

In this section we introduce a fast atom selection algorithm that reduces the number of position searched and the number of angles evaluated.

4.1. Coordinates selection
In order to reduce this computationally demanding task, we make some assumptions about the residual image to be coded. We assume that the energy in the image is concentrated in the areas where the motion predicted model was inadequate. In particular, the points where the atoms have to be set are chosen by selecting the maximum value of the absolute difference in the motion residual frame. In the exhaustive search approach each dictionary

structure is centered at each location in the block area and the inner product between the structure and the SxS region of image data is computed.

Choosing directly the position where the atom has to be set, brings an improvement of a factor S^2 in the number of times the inner product has to be computed.

4.2. Angle selection

Atoms are identified by position, scale factors and rotation. We propose an algorithm to reduce the operations needed to compute the exact rotation for the atom. The full search procedure tests every angle from the smallest to the largest, selecting the one whose projection has the highest value. Let's remind that $\theta \in [0, \pi[$ and we express it using an integer n such that $\theta = \dfrac{n \cdot \pi}{128}$.

The assumption we made is that the value of the scalar product between the residual image and the atom increases as soon as we are getting closer to the right angle, for most of the atoms. We propose a method based on a dichotomist search.

First the algorithm tests four angles set at $i\delta$, with $i=[0..3]$ and $\delta = \Pi/4$; these are called respectively A,B,C,D in Fig.1. Once found the best matching angle, then a dichotomist process starts which keeps on dividing by two the angle until we get to the unit angle $\gamma = \Pi/128$. At each step two comparisons are made and the rotation angle for which the best angle is found becomes the angle around which the new rotation angles are searched.

Indicating with N=128 the possible angles to be searched, then the complexity Nr of the proposed search can be expressed as follows:

$$Nr = 4 + 2\log_2 \delta / \gamma . \tag{11}$$

The gain G in terms of complexity of the proposed approach when compared to the full search is the following.

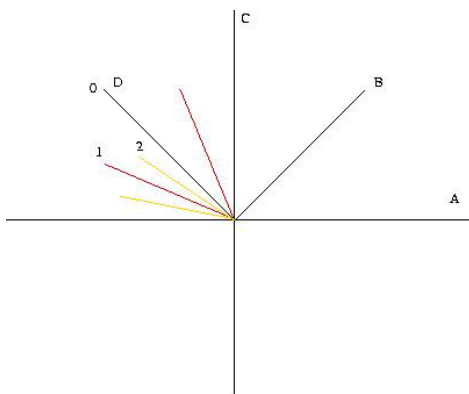$$G = \frac{S^2 N}{4 + 2\log_2\left(\dfrac{\delta}{\gamma}\right)} \tag{12}$$

Fig. 1 Dichotomist algorithm for the angle selection. First four positions checked are A,B,C,D. 1 and 2 represent the first two steps of the log search.

5. EXPERIMENTAL RESULTS

In our experiments we have used the two functions g_0 and g_1 and we have compared the different search strategies. FS is the Full Search; Max+FS is the algorithm that places the atom at the Max of the MAD of the dfd frame and uses a fullsearch for the 3 other parameters. GA is a genetic algorithm as described in [7]. Max+GA sets the atom adopting the same approach as the previous described method and uses a GA for the other parameters. Max+ang sets the atoms as the previous two methods and adopts the angle selection method described in subsection 4.2; the scale parameters are searched using an exhaustive approach.

We have used these different search algorithms with the dictionary obtained from the generating functions g_0 and g_1. Results in table 1 have been obtained coding the sequence Stefan. We can see that the function g_0 shows a better behavior than g_1. Results are expressed in MSE and number of atoms needed to code the frame.

The spread between the two dictionaries increases when sub-optimal search algorithms are used. Probably this is due to the fact that the peaky generating function g_0 suits particularly well when the atom is set in the position where there is a peak in the residual.

We can notice that the criterion of setting the atom in the Max of the absolute value introduces a 10% increase in MSE when compared to an exhaustive approach.

On the contrary, the same criterion (Max+GA algorithm) introduces a 10% improvement in the MSE when compared to a complete GA algorithm.

Results for the Max+ang. algorithm show that the idea that the MSE converges as soon as the atom gets closer to the right orientation is quite correct. In fact the drop in MSE, when compared to the Max+FS approach is very limited. We can state that the logarithmic search doesn't influence so much the quality performance.

On the contrary a big impact on the final result is given by the position chosen to set the atom.

	MSE g0	MSE g1	Atoms g0	Atoms g1
FS	193	198	382	382
Max+FS	218	230	387	389
Max+GA	223	236	393	395
GA	249	256	401	404
Max+ang.	221	229	391	388

Table 1 Comparison of the two dictionaries derived from g0 and g1 and of the various atom selection algorithms for the sequence Stefan

Concerning the complexity, we have a great improvement over the full search which is expressed in Eq. (12), while because of the difficulties to measure the convergence of the GA we can in absolute say that implementations of the

Max+ang. and GA takes approximately the same time to code a frame.

Fig. 2.a and 2.b are two typical examples of the $MSE = \|Rf\|^2$ behavior when coding a block with the different algorithms so far proposed. We have selected a good and a bad motion compensated block. MP stops at the 7th iteration or when MSE is 65. FS is the limit to be reached. Max+FS works better than GA.

Another interesting observation that came out of our results is that the parameters of the atoms for the dictionary obtained from g_0 and g_1 are quite similar. Position, rotation and scale factors are very close, so a possible improvement could be to use more dictionaries and try different functions once determined the atoms parameter. This would imply a limited computational impact, since just one more match would be necessary. Moreover, the cost of coding another dictionary would be just one bit per atom, while adaptive dictionaries could probably better match the evolving structure of the residual.
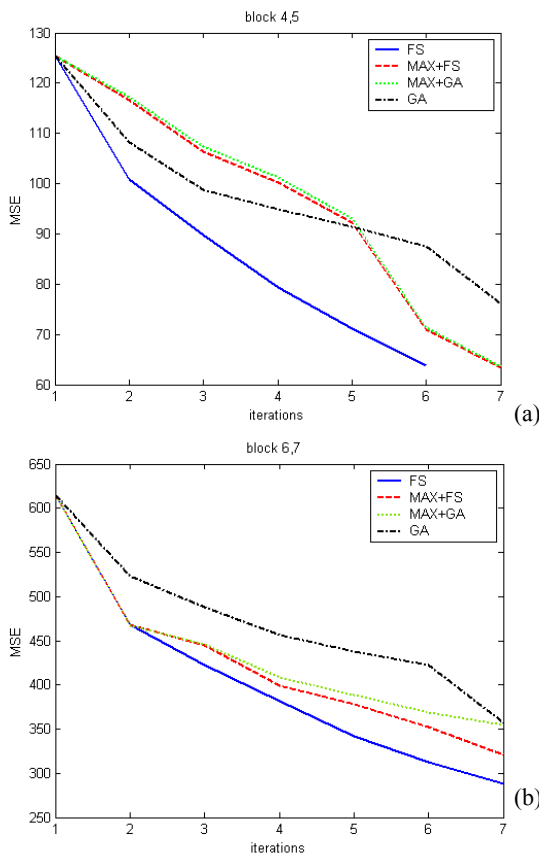
6. CONCLUSIONS

We introduced a new Matching Pursuit dictionary and a new strategy for the atom selection. The dictionary proposed shows a good behavior for dfd images.

The search algorithms introduced dramatically reduce the computational complexity when compared to the exhaustive search, while the quality impact is limited. This new methodology shows some improvements when compared to another sub-optimal strategy based on a genetic algorithm. Parameters of the two examined dictionaries are quite similar once the atom is selected, which suggests that adapting dictionaries could probably better represent the evolving nature of the residual. Variable dictionaries might indeed improve the coding efficiency and represent an interesting topic for further investigations.

7. REFERENCES

1. Neff R. and Zakhor A., *Very Low Bit-Rate Video Coding Based on Matching Pursuit*. IEEE Trans. Circuits Syst. Video Technol., February 1997. **vol. 7**(no. 1): p. 158-171.
2. Osama K. Al-Shaykh, et al., *Video Compression Using Matching Pursuit*. IEEE Trans. Circuits Syst. Video Technol., February 1999. **vol. 9**(n.1): p. 123-143.
3. Vandergheynst P. and Frossard P. *Efficient Image Representation by Anisotropic Refinement in Matching Pursuit*. in *ICASSP 2001*. 2001. Salt Lake City.
4. Mallat S. and Zhang Z., *Matching Pursuit with Time-Frequency Dictionary*. IEEE Trans. on Signal Processing, December 1993. **vol. 41**(n. 12): p. 3397-3415.
5. Frossard P. and Vandergheynst P. *Redundancy in Non-Orthogonal Transforms*. in *ISIT*. 2001. Washington DC.
6. L. Granai, Master Thesis *Codifica Video con Matching Pursuit*. October 2001, University of Siena, Italy.
7. Rosa M. Figueras y Ventura, Pierre Vandergheynst, and P. Frossard. *Evolutionary Multiresolution Matching Pursuit and its Relations with Human Visual System*. to appear in Proceedings of *Eusipco*. 2002.

Fig. 2. MSE versus number of iteration in two blocks of the sequence Stefan