16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

# INTERLEAVED S+P SCALABLE CODING WITH INTER-COEFFICIENT CLASSIFICATION METHODS

*François Pasteau, Marie Babel, Olivier Déforges, Laurent Bédat*

IETR/Image group Lab
CNRS UMR 6164/INSA Rennes
20, avenue des Buttes de Coësmes
35043 RENNES Cedex, France
fpasteau@ens.insa-rennes.fr {mbabel, odeforge, lbedat}@insa-rennes.fr

## ABSTRACT

Next generations of still image codecs should not only have to be efficient in terms of compression ratio, but also propose other functionalities such as scalability, lossy and lossless abilities, region of interest coding, etc. In previous works, we have proposed the LAR compression method covering these requirements. In particular, the Interleaved S+P scheme offers an efficient mean to compress images, especially in the medical field. In this paper, two classification methods are proposed in order to increase the compression ratio of the coder. The first one is based on a spatial context, the second one takes into account the local activity of the picture. Results are then discussed and compared to the state of the art, thus revealing high compression performances of our coding solution.

## 1. INTRODUCTION

Despite many drawbacks and limitations, JPEG is still the most commonly-used compression format in the world. JPEG2000 overcomes this old technique, particularly at low bit rates, but at the expense of a significant complexity overhead. Therefore, the JPEG normalization group has recently proposed a call for proposals on JPEG-AIC (Advanced Image Coding) in order to look for new solutions for still image coding techniques [1]. Its requirements reflect the earlier ideas of Amir Said [2] for a good image coder : compression efficiency, scalability, good quality at low bit rates, flexibility and adaptability, rate and quality control, algorithm unicity (with/without losses), reduced complexity, error robustness (for instance in a wireless transmission context) and region of interest decoding at decoder level. A. Said and the JPEG committee should add additional functionalities such as image processing at region level, both at the coder or the decoder. The LAR (Locally Adaptive Resolution) tries to address all these features. In [3], we proposed an original scheme able to perform efficient lossy compression, enabling an unusual hierarchical region representation (without any shape description). Then, in [4], we presented an extension of a more efficient scalable multi-resolution solution in terms of both lossy and lossless compression, the LAR Interleaved S+P. New improvements have been shown by the introduction of the Reversible Walsh Hadamard Transform (RWHaT) in [5].

The work presented here was based on the Interleaved S+P scheme and has been designed to produce a better compression ratio especially in lossless coding.

It is indeed known that, in some fields such as medicine, image compression must be losslessly realized to prevent a misdiagnosis [6]. Moreover, laws force hospitals to store all their pictures during at least 30 years which leads to a big amount of produced and stored data each year. Another critical application domain concerns cultural digital libraries [7]. Museums actually try to safely digitalize their belongings and thus produce large quantities of lossless picture. In France, the TSAR project has been created to develop an efficient mean to compress and secure high resolution images with the collaboration of the Louvre museum [8].

The use of the Interleaved S+P coding scheme has been motivated by the fact that this coder has been proven to be really efficient on medical images with far better results than the state of the art. It also offers some interesting features such as scalability or Quadtree partitioning. However, the compression efficiency can be greatly increased by the mean of efficient coefficient classification [9]. To propose efficient classification methods well adapted to the Interleaved S+P codec, two different approaches of the problem have been studied resulting in two different schemes, each of them leading to particular features.

The paper is organized as follows. The following section introduces the basic of the LAR coder and the Interleaved S+P scheme. In section 3, the first method of context based classification method is presented followed in section 4 by the second method. In the section 5, results of the different schemes are shown and discussed.

## 2. LAR INTERLEAVED S+P

### 2.1 LAR overview

The basic concept of the LAR method is that local resolution should be adapted to suit local activity. Also assuming that an image consists of global information and local texture, we firstly proposed a two-layer, content-based codec, both relying on a quadtree partition. The first layer, called the FLAT LAR, encodes the global information at block level representation. The additional second layer enables texture compression within blocks. Therefore, the method provides natural SNR scalability. The block sizes are estimated through a local morphological gradient. The direct consequence is that the smallest blocks are located round the edges whereas large blocks map homogeneous areas. This being so, the main feature of the FLAT coder consists of preserving contours while smoothing homogeneous parts of the image. This characteristic is also exploited to get a free hierarchical region representation : from the low bit-rate image compressed by the FLAT LAR, both coder and decoder can perform a segmentation process by iteratively merging blocks into regions. A direct application is then Region Of Interest (ROI) enhancement, by first selecting regions and enabling second layer coding only for the relevant blocks.

In order to obtain higher image quality, the texture (whole error image) can be encoded by the second layer called spectral coder that uses a DCT adaptive block-size approach. The use of adapted square size allows a content-based scalable encoding scheme : for example, edge enhancement can be made by only transmitting the AC coefficients of small blocks.

### 2.2 The Interleaved S+P scheme

To perform the decorrelation of the picture, the interleaved S+P scheme is used. The S+P transform (S-transform + Prediction) is based on the 1D S-transform applied on the 2 vectors formed by 2 diagonally adjacent pixels in a 2x2 block as depicted in figure 1. Let $z_0$ and $z_1$ denote the S-transformed coefficients and $(u_0, u_1)$ be the couple of values, we have :

$$\left| \begin{array}{l} z_0 = \lfloor (u_0 + u_1)/2 \rfloor, \\ z_1 = u_1 - u_0. \end{array} \right. \qquad (1)$$

The prediction is achieved in 3 successive passes. If $i \in \{0, 1\}$ and $k \in \{1, 2, 3\}$, $z_i^k$ consitutes the $z_i$ coefficient coded through the $k^{th}$ pass. Let I be the original image of size $N_x \times N_y$. The multiresolution representation of an image is described by the set $Y_{ll=0}^{lmax}$, where $l_{max}$ is the top of the pyramid and l = 0 the full resolution image. Four blocks $\frac{N}{2} \times \frac{N}{2}$ are gathered into one block $N \times N$ valued by the average of the two blocks of the first diagonal (first S-pyramid on fig. 2) The transformation of the second diagonal of a given $2 \times 2$ block can also be seen as a second S-pyramid, where the pixel values depend on the ones existing at the lower level of the first S-pyramid. Interleaving is in this way realized.

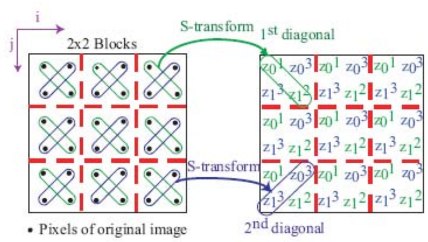



FIG. 1 – S-Transform scheme

When using the Interleaved S+P scheme, $z_0^1$ coefficients of a given layer are automatically retrieved from the upper layer. That is why only three types of coefficients $z_1^2$, $z_0^3$ and $z_1^3$ have to be estimated for each level. This estimation leads to three different types of errors, each corresponding to one type of coefficients. The first layer coding (FLAT LAR) builds the first pass of the pyramid used by the Interleaved S+P. It decomposes each pixel of a given layer into a $2 \times 2$ block into a lower level according to the information given by the quadtree. To perform a lossless compression, the second layer coding (texture) performs a second pass on this pyramid. It decomposes every pixels that have not been decomposed previously.

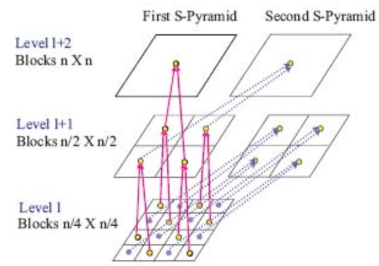



FIG. 2 – S pyramid scheme

As $z_1^2$, $z_0^3$ and $z_1^3$ coefficients present different characteristics in terms of errors distribution, the entropic classification schemes are performed independently for each type of errors. It leads to three different types of classification for both FLAT LAR and Texture. The aim of context modeling is to perform discriminations inside each type of errors. Thus, coefficients of a given type producing the same error distributions are gathered in one context class.

## 3. INTER AND INTRA-LEVEL CONTEXT MODELING

In this section, a context modeling technique is designed, considering error values in the neighborhood. It seems natural that when the predictor makes large errors on the neighbors of a coefficient, the error on this coefficient is most likely large. This way, inter-level and intra-level contexts can be taken into account. Inter-level context refers to the context represented by the neighbors of higher levels whereas intra-level context refers to the context represented by the neighbors of the same level.

### 3.1 Intra-level and inter-level context

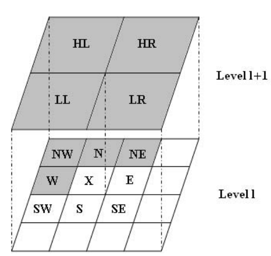


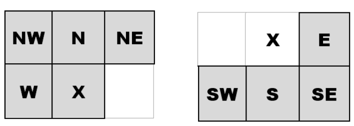

FIG. 3 – "360 + 1" degree context

Due to the raster image scan, the causal neighborhood of a coefficient, in intra-level context, is only composed of four neighbors W, N, NW and NE as described in the figure 3. This neighborhood represents a 180 degree context around the error to be coded.

The quality of the context can be increased by adding information coming from higher layer leading to consider an inter-level context. This way, it is possible to add to the remaining intra-level context, which represents the most reliable information, some neighbors such as E, S, SW, and SE (figure 3). Moreover, it is now possible to use the value of the error (here HL) directly located above the error to be coded.

This leads to an extended 360 degree context composed of 9 neighbors that can be naturally split into two homogeneous groups in terms of error correlation. Moreover, this enhanced context manages to work out the problem of the "holes" created by the quadtree partitioning and conditional decomposition during the first pyramid pass. In fact, the error made on the block of the last layer which has been decomposed is copied in these "holes", and therefore leading to a better adaptation to the context. As a consequence, the error located at the position X when the classification is performed corresponds to the error located directly above (here HL).

### 3.2 Threshold estimation




Group $G_1$ Group $G_2$

FIG. 4 – Set of position $G_1$ and $G_2$

Let $E_l$ be the error image at level $l$. We note $(E_l(i,j))_p$ as the translation of $E_l$ at position $(i,j)$ by $p$. $G_1$ is defined as the set of positions $\{NW, N, NE, W, X\}$ and $G_2$ as $\{SW, S, SE, E\}$ (figure 4).

We defined :

$$\begin{aligned} E_{MG1}(i,j) = max\{(E_l(i,j))_p | p \in G_1\}, \\ E_{MG2}(i,j) = max\{(E_l(i,j))_p | p \in G_2\}. \end{aligned} \quad (2)$$

We also define $T_k^{G_1}$ and $T_k^{G_2}$ for $k \in \{1..NbClass\}$ so that :

$$\left| \begin{array}{l} \text{if } (E_{MG1}(i,j) < T_1^{G_1} \text{ and } E_{MG2}(i,j) < T_1^{G_2}) \\ \quad \text{then } E_l(i,j) \in Class_1 \\ \text{elseif } (E_{MG1}(i,j) < T_2^{G_1} \text{ and } E_{MG2}(i,j) < T_2^{G_2}) \\ \quad \text{then } E_l(i,j) \in Class_1 \\ \ldots \\ \text{else } X \in Class_{Nbclass}. \end{array} \right. \quad (3)$$

Because the predictor is not biased, only the absolute part of the value of the neighbors is used to perform the classification. As stated before, the value of the neighbors N, W, NW and NE are the most reliable because they belong to the intra-level context. The neighbor noted X gives us an interesting information on the inter-level behavior of the error ($G_1$). Therefore they are tested with lower thresholds with $T_k^{G_1} \leq T_k^{G_2}$. To prevent a context dilution [9], we perform a context quantization with only three classes. The values of the thresholds are determined by experimentation and learning. Threshold values estimated from a restricted set of pictures are revelant even for the same type of images.

## 4. LOCAL ACTIVITY CLASSIFICATION

The classification method presented here is based on the one used in the RWhaT+P Classification [5]. This classification is really efficient in the case of the RWHaT decorrelation scheme. The RWHaT and Interleaved S+P methods share the same global decomposition scheme (2 pass pyramidal decomposition). The main difference relies in the elementary decomposition steps, performed by a reversible form of the common $2 \times 2$ blocks WHT for the RWHaT. The objective is to adapt the concept of inter-coefficient classification technique proposed for the RWHaT to the Interleaved S+P.

### 4.1 Local Activity

This classification scheme is based on the *a priori* determination of the local activity of 2x2 blocks. Because of the differences between the RWHaT Transformation and S-Transformation, the RWHaT calculation method cannot be directly applied. A new determination of activity is proposed. Both the prediction of pixels of the current block and the value of the pixel from the higher level are used for this purpose (figure 5). The differences between the predictions and the value of the pixel from the higher level are calculated before applying S-transforms. Four coefficients $\tilde{z}_0^1, \tilde{z}_1^2, \tilde{z}_0^3, \tilde{z}_1^3$ are defined by :

$$\left| \begin{array}{l} \tilde{z}_0^1 = [(HL - B) + (LR - B)]/2, \\ \tilde{z}_1^2 = (HL - B) - (LR - B), \\ \tilde{z}_0^3 = [(HR - B) + (LL - B)]/2, \\ \tilde{z}_1^3 = (HR - B) - (LL - B). \end{array} \right. \quad (4)$$

Each of them represents the contribution of the coefficient, respectively $z_0^1, z_1^2, z_0^3, z_1^3$, to the gap of energy observed between two levels. To increase the efficiency of the classification, the flat coefficients are discriminated from the texture ones. In case of flat coefficients, values of $\tilde{z}_1^2, \tilde{z}_0^3, \tilde{z}_1^3$ are directly used in the classification process as respectively $A_1$, $A_2$ and $A_3$.

For texture coefficients, only two activities $A_1$, $A_2$ are computed in such a way that :

$$\left| \begin{array}{l} A_1 = \tilde{z}_1^2, \\ A_2 = \tilde{z}_0^3 \times \tilde{z}_1^3. \end{array} \right. \quad (5)$$

The first activity $A_1$ corresponds to the activity of the coefficient $z_1^2$ and $A_2$ to the activity of the coefficients $z_0^3, z_1^3$.

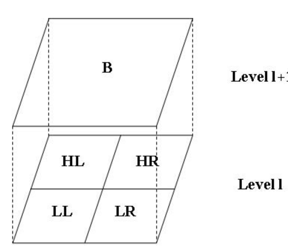



FIG. 5 – Local Activity

### 4.2 Coefficient classification

The coefficient classification consists of gathering the coefficients into equiprobable classes.

Let $Nb_{class}$ be the maximal number of classes. Let $p_A$ be the probability : $p_A = p(A = A_i)$ with $1 \leq i \leq 3$ for flat coefficients and $1 \leq i \leq 2$ for texture ones. Let $d_n$ be such a way that if $d_0 = \infty$ and $d_{Nb_{class}} = 0$, we can write :

$$\forall d_n \in \{d_1, \ldots, d_{Nb_{class}}\}, \quad \int_{A=d_{n-1}}^{d_n} p_A = \frac{1}{Nb_{class}}. \tag{6}$$

the definition of a class $C_n$ is given by :

$$C_n = \{A | d_{n-1} > A \geq d_n\}, \tag{7}$$

with $1 \leq n \leq Nb_{class}$.

Each class constitutes a specific substream sent to the entropy coding layer. The maximal number of classes $Nb_{class}$ is defined for each level independently : 4 for the full resolution, 3 for level 1, and so on ...

Level 2 Level 1 Level 0

FIG. 6 – Local Activity picture

### 4.3 Integration of neighbor activities

By definition (equation 4), the local activity only uses the values of the current block and the one from the higher level. Thus the quality of this classification can be increased by taking into account the neighbors activities. The scheme used here is really similar to the one implemented in the inter and intra-level context modeling. The computation of the activities is made in a first pass on the whole image. As a consequence, the context around the block contains the 8 neighbors of this block. A 360 degree context is thus created. Then these neighbor activities are compared to predetermined thresholds. Empirically, it has been noticed that a number of classes higher than 4 does not increase the compression ratio. Therefore, the number of thresholds has been set to 3, which leads to 4 different classes.

Combining these two schemes produces a new adaptative classification different for each layer, leading to design 4 classes for higher level and 16 classes for highest resolution.

## 5. RESULTS

In this section, the results of different tests are shown and analysed. A set of 90 medical images and 12 natural images have been chosen in order to perform those tests.

The results presented in table 1 have been obtained when encoding natural images. The thresholds required for the classification of the inter and intra-level context modeling have been empirically determined.

Results relative to natural images are quite heteroclite but performing context modeling is always benefic to the compression ratio. Generally speaking, we can see that the results obtained thanks to the local activity scheme are better than the ones obtained by the Inter/Intra Level one with an average gain of 0.05 bpp and far better than without coefficient classification with an average gain of 0.134 bpp. The most noticeable increase of compression ratio can be observed on us-image with an overall gain of 0.38 bpp. In comparison with the state of the art coding scheme CALIC, the coder presented here produces an average gain of 0.152 bpp. Once again, the best improvement of compression ratio is realized on us-image and is equal to 0.59 bpp.

Table 2 shows entropic costs with the different schemes previously described obtained on medical images such as mammograms. The thresholds for inter and intra-level context modeling have been empirically determined on a restricted set of 5 pictures and then directly applied to the whole set.

Unlike results observed on natural images, coefficient classification with medical images like mammograms produces steady results. In most cases, inter and intra-level context modeling leads to a better compression ratio than the local activity classification. The average gain is about 0.1 bpp in comparaison with the Interleaved S+P alone, and about 0.406 bpp with CALIC.

Even if tests have been realised on a set of 90 images, only a subset is shown here wich represents the trends generally observed on all the images. The table 3 shows the relative gain in percentage between the results obtained with predetermined thresholds and the ones with optimal thresholds for inter and intra-level context modeling with medical images. The average relative gain is 0.30% which represents 0.004 bpp. In comparaison with the reduction of 0.092 bpp in regard of Interleaved S+P alone offered by this scheme, the gain is quite small. This confirms that these thresholds are steady for images of the same kind and that their determination can be done offline on a reduced set of image before being used on the rest of the set.

## 6. CONCLUSION

The Interleaved S+P scheme leads to interesting results, especially in the medical field, but it can be greatly improved by the mean of an efficient coefficient classification method. In this paper, two different ways of realizing this classification have been presented. Considering the inter and intra-level context modeling technique, it appears that this kind of classification highly improves the compression ratio of medical images but also need the determination of certain thresholds to achieve such results. The local activity classification leads to far better results than the previous method in the field of natural images and without any special threshold predetermination. However, results on medical images remain below the one produced by the first method.

These two classification schemes appear to be complementary : Inter/Intra Level context modeling is really efficient on a given set of images as soon as the thre-

| Entropy (bpp) | | | | | |
|---|---|---|---|---|---|
| Image | Raw | CALIC | Interleaved S+P | Inter/Intra Level | Local Activity |
| Barbara2 | 7.51 | **4.93** | 5.06 | 4.98 | 4.96 |
| Hotel | 7.57 | **4.57** | 4.76 | 4.71 | 4.61 |
| Lena | 7.44 | 4.33 | 4.31 | **4.22** | **4.22** |
| Peppers | 7.57 | 4.58 | 4.60 | 4.53 | **4.45** |
| tools | 7.52 | 5.53 | 5.57 | 5.48 | **5.44** |
| us | 4.84 | 3.60 | 3.39 | 3.12 | **3.01** |
| Average | 7.150 | 4.600 | 4.615 | 4.506 | **4.448** |

TAB. 1 – $zero^{th}$ order entropy results on natural images

| Entropy (bpp) | | | | | |
|---|---|---|---|---|---|
| Image | S+P | CALIC | Interleaved S+P | Inter/Intra Level | Local Activity |
| mdb006 | 2.21 | 2.14 | 1.78 | **1.72** | **1.72** |
| mdb009 | 2.07 | 1.96 | 1.82 | **1.63** | 1.68 |
| mdb011 | 1.96 | 1.88 | 1.68 | **1.49** | 1.56 |
| mdb070 | 2.13 | 2.04 | 1.78 | **1.65** | 1.71 |
| mdb076 | 1.82 | 1.75 | 1.44 | **1.32** | 1.33 |
| Average | 2.038 | 1.954 | 1.700 | 1.562 | 1.600 |

http ://peipa.essex.ac.uk/ipa/pix/mias/

TAB. 2 – $zero^{th}$ order entropy results on medical images : Mammograms

| Image | Gain(%) | Image | Gain(%) | Image | Gain(%) |
|---|---|---|---|---|---|
| mdb006 | 0.07 | mdb005 | 0.09 | mdb028 | 0.17 |
| mdb009 | 0.32 | mdb010 | 0.40 | mdb075 | 0.12 |
| mdb011 | 0.02 | mdb012 | 0.17 | mdb092 | 0.65 |
| mdb026 | 0.16 | mdb025 | 0.07 | mdb095 | 0.16 |
| mdb060 | 0.72 | mdb059 | 0.11 | mdb134 | 0.50 |
| mdb070 | 0.10 | mdb069 | 0.16 | mdb141 | 0.34 |
| mdb076 | 0.65 | mdb080 | 0.65 | mdb144 | 0.62 |
| mdb077 | 0.07 | mdb091 | 0.12 | mdb148 | 0.64 |
| AVERAGE | 0.24 | AVERAGE | 0.23 | AVERAGE | 0.36 |

http ://peipa.essex.ac.uk/ipa/pix/mias/

TAB. 3 – Mammograms - Influence of context determination thresholds

sholds have been optimized. However Local Activity classification leads to good results on natural images directly out of the box.

Even if these thresholds Inter/Intra Level context modeling are quite steady, a further study has to be done so that to automatize their determination.

**Acknowledgment :** This work is supported by the French National Research Agency as part of the TSAR project.

## REFERENCES


[1] Jpeg-aic : scope and evaluation. *International Standards Organization working document, ISO/IEC SC29/WG 1/N4326*, 2007.

[2] A. Said and W. Pearlman. Reversible image compression via multiresolution representation and predictive coding. In *Visual Communication and Image Processing*, volume 209, pages 664–674. SPIE, Novembre 1993.

[3] O. Déforges, M. Babel, L. Bédat, and J. Ronsin. Color LAR Codec : A Color Image Representation and Compression Scheme Based on Local Resolution Adjustment and Self-Extracting Region Representation. *IEEE Trans. on Circuits and Systems for Video Technology*, 17(8) :974–987, August 2007.

[4] M. Babel, O. Deforges, and J. Ronsin. Interleaved S+P Pyramidal Decomposition with Refined Prediction Model. In *ICIP*, volume 2, pages 750–753, October 2005.

[5] O. Déforges, M. Babel, and J. Motsch. The RWHT+P for an improved lossless multiresolution coding. In *EUropean SIgnal Processing COnference, EUSIPCO'06*, Florence, Italy, September 2006.

[6] P. White. Legal issues in teleradioogy - distant thoughts ! *The British Journal of radiology*, 75 :201–206, 2003.

[7] G. F. MacDonald. Digital visionary. *Museum News*, 79 :34–41, March/April 2000.

[8] M. Babel, L. Bédat, O. Déforges, and J. Motsch. Context-Based Scalable Coding and Representation of High Resolution Art Pictures for Remote Data Access. In *Proc. of the IEEE International Conference on Multimedia and Expo, ICME'07*, pages 460–463, July 2007. Projet ANR TSAR.

[9] X. Wu. *The Transform and Data Compression Handbook*, chapter Compression of Wavelet Transform Coefficients. K. R. Rao et al., Boca Ratin, CRC Press LLC, 2001.