A Preprocessing Technique for Improving the Compression Performance of JPEG 2000 for Images With Sparse or Locally Sparse Histograms

Souha Jallouli*, Sonia Zouari*, Atef Masmoudi*†, William Puech†, Nouri Masmoudi*
* University of Sfax, Laboratory of Electronics and Technologie of Information,
National Engineering School of Sfax, 3038 Sfax, Tunisia
† University of Montpellier, LIRMM, UMR CNRS 5506, 34392 Montpellier Cedex 05, France
souhajallouli@enis.tn
sonia.zouarikrichen@gmail.com
masatef@gmail.com
william.puech@lirmm.fr
masmoudi123@gmail.com

***Abstract*—JPEG 2000 is one of the most efficient and well performing standards for continuous-tone natural images compression. However, a compression performance loss may occur when encoding images with sparse or locally sparse histograms. Images of the later type include only a subset of the available intensity values implied by the nominal alphabet. This article proposes a new adaptive block-based histogram packing which improves the lossless compression performance of JPEG 2000 with sparse histogram images. We take advantage, in this work, of the strength likelihood between symbol sets of the neighboring image blocks and the efficiency of the offline histogram packing with sparse or locally sparse histogram images. Results of its effectiveness with JPEG 2000 are presented.**

***Index Terms*—lossless image compression; offline histogram packing; JPEG 2000; alphabet reduction scheme; sparse and locally sparse histograms.**

I. Introduction

Different types of image content have been displayed in the few pass decades, e.g., images may contain textual, graphical or computer generated items besides the natural content. Frequently, for this kind of images, the actual number of active symbols, which correspond to the pixels' intensities, is lower than implied by the nominal bit depth. In this case, the active intensities spread all over the nominal intensity range and not all the bins in a histogram are used. The histogram is therefore sparse and we call such images as having sparse histogram. The high dynamic range, medical and astronomical images tend to have sparse histograms since they have longer bit depth than a current standard image [1]. An image pre-processing may also be the reason of image histogram sparseness such as a histogram modification, a tone mapping, a Gamma correction and an extraction of a region of interest [2].

Despite the phenomenal success of its algorithm based on wavelet transform, JPEG 2000 has several shortcomings that become increasingly apparent. JPEG 2000 is among the most recent ISO/ITU standards which were designed mainly to compress continuous-tone natural images [3]–[6]. However, its compression efficiency is severly affected when handling images with sparse or locally sparse histograms [7]. Histogram sparseness is clearly different from what is expected by transformation image compression algorithms. This is can be due to the statistical model distortion caused by the zero frequency problem [8] when using an entropy coding technique.

In literature, there are several proposal of approaches that modify the coding of JPEG 2000 for lossless compression of images with sparse or locally sparse histograms [9], [10]. The aim of this work is to improve the compression performence of JPEG 2000 for this kind of images using a preprocessing technique without any modification of its code.

II. Histogram Packing

The impact of histogram sparseness on image compression ratios is well known [11], [12]. Recent studies show that the compression efficiency of JPEG 2000 can be improved by the use of histogram packing [13].

The simplest method, among these preprocessing techniques, is called the offline histogram packing [7], [14]. It simply maps all the intensity levels into a contiguous set corresponding to the lowest part of the nominal intensity range. The sparse histogram is therefore transformed on a dense one. To be reversible, this transform requires the original histogram to be encoded along with the compressed image. This approach, when applied prior to JPEG 2000, supply an important improvement. However, the global offline considers intensity values that appear a few number of times in the image as having equal importance as those that occur most frequently. Thus, images with quasi−sparse histograms can not benefit from this technique.

A further interesting alternative to the global histogram packing is known as Block-based histogram packing [15]. In fact, a histogram packing can be further improved if the histograms of image blocks are packed instead of the golbal one. Therefore, the basic idea was to apply the histogram

ISBN 978-0-9928626-7-1 © EURASIP 2017

packing on consecutive image blocks which provides the ability to explore the non stationary local characteristics. Several studies have confirmed that encoding on block-basis provide additional improvements particularly for images with locally sparse histograms [16]–[18]. However, the increase in the number of blocks leads to a compression efficiency loss since the total overhead generated by block-based histogram packing methods depends on the number of image blocks and the size of the mapping table.

Recently, Masmoudi et al. [19] proposed a new compression scheme which reduces the symbol set used by Arithmetic coding to encode separatly each image block with the minimum overhead bits. With this alphabet reduction scheme (ARS), the effect of the zero frequency problem is reduced when an entropy coding technique is used.

In this paper, we propose to take advantage of the efficiency of the alphabet reduction scheme (ARS) which attributes for each block a symbol set as close as possible to its corresponding active symbol set with the minimum overhead bits. When associated with the offline histogram packing, ARS yields very efficient compression results.

III. PROPOSED METHOD

The aim of this work is to attribute for each block, a set of symbols as close as possible to its active symbols, with the minimum overhead bits as possible. For this purpose, the original image is divided into small and non-overlopping blocks and then each block is encoded separatly.

In this section we consider X^k as an image block to encode which is composed of m pixels $X^k = x_1, ..., x_m$. Each pixel x_i, with $i \in \{1, ..., m\}$, takes a value from an alphabet $\mathcal{A}$ composed of n symbols, $\mathcal{A}$ = α_1, ... , α_n. We denote by P the probability distribution of the symbols of the alphabet $\mathcal{A}$. S^k and S^0 are the sets of active symbols corresponding to the current block and the entire original image:

$$S^0 = \{\alpha_i \in S / P(\alpha_i) \neq 0\} \tag{1}$$

$$S^k = \left\{\alpha_i \in S / \exists t, X^k(t) = \alpha_i\right\} \tag{2}$$

To represent the image active symbol set, we use a binary vector V with $\mid \mathcal{A} \mid$ bits length which is calculated as follows:

$$V(i) = \begin{cases} 1, & \text{if } \alpha_i \in S^0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Let us assume that N^k, is the set of active symbol sets of the neighboring blocks of the block X^k which is calculated as follows:

$$N^k = \left\{S_0^k, S_1^k, S_2^k, S_{Min-Max}^k\right\} \tag{4}$$

where S_j^k corresponds to the set of active symbols in X_j^k the j'th neighbor block of X^k and $S_{Min-Max}^k$ is the symbol set obtained by means of the minimum and the maximum symbols present in X^k:

$$S_{Min-Max}^k = \left\{Min^k, ..., Max^k\right\} \tag{5}$$

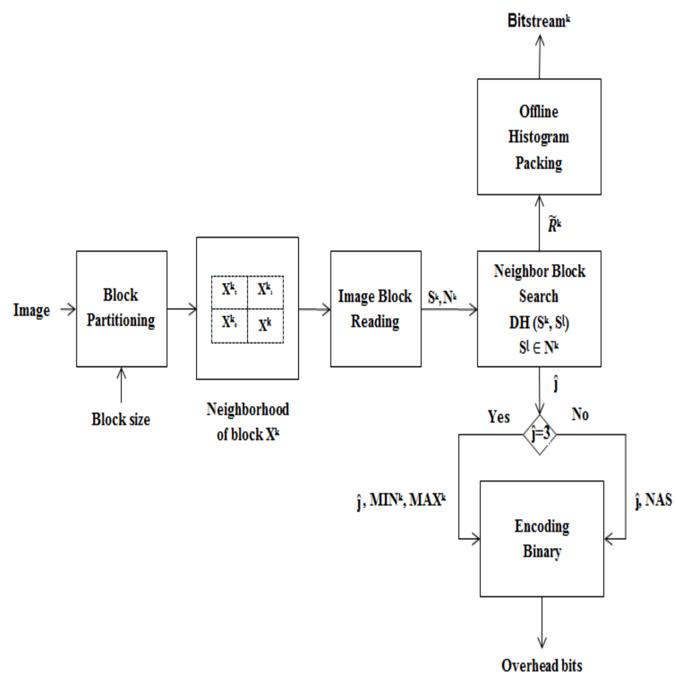


Fig. 1: Functionnal block diagram of the proposed method

We exploit, in this work, the interesting likelihood between the active symbol sets of the neighboring blocks. A major part of the active symbols in the neighboring blocks are highly likely to appear in the current active symbol set. To quantify this likelihood, the Hamming Distance is used. In fact, the Hamming Distance between two sets A and B, denoted by D_H(A,B), is the sum of the newly appearing symbols number and the disappearing symbols number [19] as below:

$$D_H(A, B) = |A \setminus B| + |B \setminus A|, \tag{6}$$

where $|.|$ denotes the set cardinality function.

A. Encoding algorithm

The block diagram of the encoding algorithm is depicted in Fig. 1 where both the ARS method and the offline histogram packing are included. According to equations (1) and (3), respectively, we calculate first the image active symbol set S^0 and a binary vector V which is then encoded as additionnal information. When encoding the current block X^k, we start by calculation of its active symbol set S^k, minimum Min^k, maximum Max^k, and the set of its neighbor block active symbol sets N^k. Next, to attribute to the current block the smallest set of active symbols, we search the closest neighbor block using the Hamming Distance. In fact, it corresponds to the neighbor block for which $D_H(S^k, S^l)$ is minimum where $S^l \in N^k$. Its position in N^k called $\hat{j}$ is then encoded as additionnal information with $\lceil log_2(3+1) \rceil$ bits.

Two scenarios are envisaged according to the value of the position $\hat{j}$:

- $\hat{j} \in \{0, 1, 2\}$: $\hat{j}$ corresponds to the position of one of the neighboring blocks.

ISBN 978-0-9928626-7-1 © EURASIP 2017

If $\hat{j}$ corresponds to the position of one of the neighboring blocks, then $\hat{S}^k$ corresponds to the $\hat{j}$-th neighbor block, the set of the newly appearing symbols NAS is therefore calculated as the difference between S^k the active symbol set of X^k and $\hat{S}^k$, $|NAS|$ is its length. To provide to the block X^k the minimum of overhead bits, we encode the index of each newly appearing symbol in $S^0 \setminus \hat{S}^k$ with $\lceil log_2(|S^0| - |S^k|) \rceil$ bits. If there are no newly appearing symbols, $|NAS| = 0$, then we output the bit "0".

$$NAS = S^k \setminus \hat{S}^k \tag{7}$$

- $\hat{j} = 3$

If $\hat{j}$ corresponds to the position of the last symbol set in N^k, which means that $\hat{S}^k = S^k_{Min-Max}$ then the Min^k and Max^k values need to be stored as additionnal information. We encode each symbol with $\lceil log_2(|S^0|) \rceil$ bits.

Finally, to generate the compressed $bitstream^k$, we apply the offline Histogram packing to the resulting symbol set $\tilde{R}^k$ given by:

$$\tilde{R}^k = S^k \cup \hat{S}^k \tag{8}$$

Let us assume that $\tilde{R}^k$ is an ascending sorted set of r different intensity values, $\tilde{R}^k = \alpha_0, .. , \alpha_{r-1}$, where $\alpha_i \in \mathcal{A}$ and $r \leq |\mathcal{A}|$. Then, the intensity mapping H, that is used to map $\tilde{R}^k$, maps ascending sorted intensity values into ascending sorted contiguous indexes:

$$H = (\alpha_0 \mapsto 0, \alpha_1 \mapsto 1, ..., \alpha_{r-1} \mapsto r-1)$$

The size of the compressed image is:

$$Compressed\_File\_Size = overheadbits + \sum_{i=1}^{N} bitstream^k \tag{9}$$

B. Decoding algorithm

For the reconstruction of the original image, we start by reading the binary vector V and the corresponding image active symbol set $\tilde{S}^0$. Next, we read $\lceil log_2(3+1) \rceil$ bits to find $\hat{j}$ the position of the closest neighbor block. If $\hat{j}$ corresponds to the position of one of the neighboring blocks, then the estimated symbol set used when encoding the K'th block, denoted by $\tilde{S}^k$ is calculated by:

$$\tilde{S}^k = NAS \cup \hat{S}^k \tag{10}$$

Otherwise, the Min^k and Max^k values are read to build $\tilde{S}^k$ which is equivalent to $S^k_{Min-Max}$, according to equation (5). Finally, we subtitute the indexes used when mapping $\tilde{S}^k$ to build the preprocessed image by their corresponding original intensity values in $\tilde{S}^k$.

IV. EXPERIMENTAL RESULTS

To assess the efficiency of the proposed method we used three sets of images. The first image set consists of eleven grayscale-converted version of a set used by Ausbeck in his PWC coder [20]. The second one regroupes five images taken from the university of waterloo Greyset2 collection [21]. This set was used to illustrate the poor perfermance of JPEG 2000 in the compression of this type of images [22]. The third set is composed of several natural images in order to evaluate the robustness of the proposed method when the input images have not sparse histograms. This set was also used in [14], [23].

We evaluate the coding improvement of the proposed approach by means of the bit rate, i.e., the total number of bits in the compressed file divided by the number of pixels in the original image.

TABLE I: Comparison between bit rates obtained by JPEG 2000 with (OFFHP column) and without (Normal column) offline histogram packing and with the new adaptive block-based Histogram Packing (ABBHP columns) for the first and the second image sets.

Image	**JPEG 2000**				
	Normal	**OFFHP**	**ABBHP**		
			8x8	16x16	32x32
Benjerry	4.04	2.63	**1.92**	2.07	2.15
Books	6.16	2.13	**1.69**	1.70	1.75
Cmpndd	2.33	2.09	1.45	**1.35**	1.36
Cmpndn	2.20	1.92	1.27	**1.18**	1.19
Gate	4.32	3.19	2.88	**2.85**	2.89
Music	5.52	2.11	**1.54**	2.14	1.90
Netscape	4.04	2.40	2.45	**2.35**	2.36
Sea_dusk	0.41	0.29	0.19	**0.16**	0.17
Sunset	3.04	2.70	2.40	**2.35**	2.43
Winaw	2.32	0.91	0.73	**0.69**	0.70
Yahoo	4.10	3.69	**2.18**	2.20	2.53
Average	3.50	2.19	**1.70**	1.73	1.77
France	2.01	2.01	**0.88**	0.93	1.05
Frog	6.25	5.25	**3.93**	4.0	4.21
Library	5.69	5.59	5.57	5.29	**5.23**
Mountain	6.69	5.42	5.73	5.35	**5.32**
Washsat	4.42	**2.23**	2.40	2.28	2.25
Average	5.01	4.10	3.70	**3.57**	3.61
Bike	5.06	**5.02**	7.27	5.88	5.33
Bike3	**3.76**	**3.76**	5.16	4.35	3.98
Cafe	**5.31**	**5.31**	8.17	6.16	5.48
Goldhill2	**4.84**	**4.84**	6.69	5.32	4.96
Lena2	**4.31**	**4.31**	6.15	5.00	4.56
Woman	**4.28**	**4.28**	5.71	4.70	4.44
Average	**4.59**	**4.59**	6.52	5.23	4.79

Note : OFFHP, Offline Histogram Packing; ABBHP, Adaptive Block Based Histogram Packing.

Table I presents the compression results obtained with the lossless JPEG 2000 for different sizes of blocks: 8x8, 16x16, 32x32. The analysis of the results shows clearly how the proposed method "ABBHP" can improve the losless compression

ISBN 978-0-9928626-7-1 © EURASIP 2017

of sparse histogram images, which is the case of the images of the first and the second sets. As can be observed, the proposed method ABBHP provides an interesting improvement regarding to the lossless image compression standard JPEG 2000 with and without offline histogram packing. The new approach outperforms JPEG 2000 when it is applied directly to images with a bit rate gain superior to 50% for almost the test images of the first set. Moreover, an interesting improvement is obtained even when applying offline histogram packing prior to JPEG 2000. Offline histogram packing with JPEG 2000 is slightly better for only one image, washsat, from the second set which is due to the overhead bits required for later recovery of the original intensity values. Moreover, the degradation in the compression rates of natural images (those of the third set) is due to the lack of characteristics usually found for sparse histogram images. The proposed method is specifically designed for sparse histogram images.

Table II summarizes the average bit rate reduction achieved by ABBHP relative to that of JPEG 2000 with and without offline histogram packing for the two image sets and for different block sizes. For the first image set, ABBHP achieves an interesting average bit reduction by 51.4% when compared with JPEG 2000, and 22.3% lower than that of JPEG 2000 with offline histogram packing when applied to blocks of size 8x8. In addition, ABBHP reduces the bit rate of th second set by 28.7% and 12.9% relative to that of JPEG 2000 with and without offline histogram packing, respectively.

In addition, from the observation of the rows of Table III, we can confirm that ABBHP provides an dditionnal information as small as possible. For the first set, the novel method achieves an average additionnal information equal to 5.09% when applied to blocks of size 16x16.

TABLE II: Average bit rate reduction (in %) achieved by our ABBHP relative to that of JPEG 2000 with and without offline histogram packing for different block sizes for the first and second sets.

JPEG-2000	**First Set**			**Second Set**		
	Block sizes			**Block sizes**		
	8x8	**16x16**	**32x32**	**8x8**	**16x16**	**32x32**
Normal	51.4	48.2	49.4	26.1	28.7	27.9
OFFHP	22.3	17.3	19.1	9.7	12.9	11.9

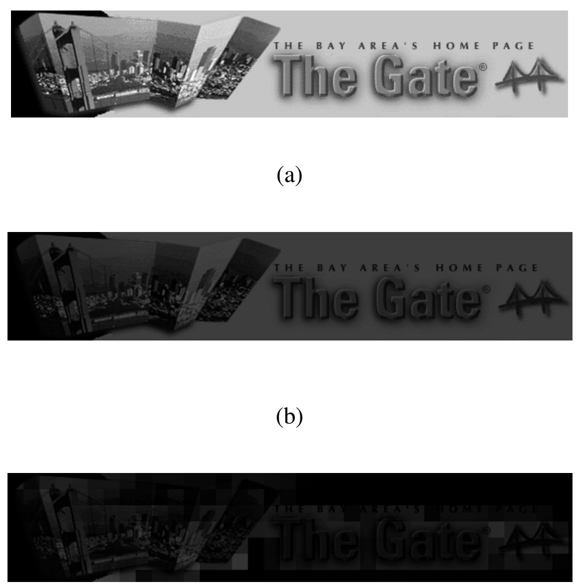
Figure 2 illustrates a gray-scale image, ”Gate”, and transformed images obtained with OFFHP and ABBHP. As can be observed, the ABBHP makes the image too dark which means that the histogram becomes more dense and centred on ”Zero” better than OFFHP. These results prove the efficiency of the proposed method for images with sparse or locally sparse histograms regarding to JPEG 2000 which represent a poor performance, even when combined with an histogram packing technique.

Note that in this work, we have not presented any implementation details for JPEG 2000 as our proposed method does not imply any modification of its code. Moreover, the proposed

TABLE III: Average additional information (in %) achieved by the proposed method for 16x16 blocks for the first set.

Image	**CFS (Bytes)**		**AI (%)**
	Bitstream	**AI**	
Benjerry	55696	2352	4.05
Books	95648	1144	1.81
Cmpndd	497816	33936	6.38
Cmpndn	432904	32016	6.89
Gate	165528	834	4.80
Music	20584	528	2.50
Netscape	138792	4896	3.40
Sea_dusk	22488	2880	11.35
Sunset	672792	48936	6.15
Winaw	200312	5152	2.50
Yahoo	53992	3536	6.15
Average	**205 134.72**	**12 382.72**	**5.09**

Note : CFS, Compressed File Size; AI, Additionnal Information.



(c)

Fig. 2: **a**) The ”Gate” image, **b**) Packed image using offline histogram packing, **c**) Packed image using the proposed method.

method was designed to be efficient and suitable for both hardware and software implementations. The coding performance of our proposed algorithm is also shown in terms of speed processing and complexity due to the restriction of the number of symbols for each block, as the number of computations is linear proportional with the size of the set of symbols. The overall complexity of the proposed algorithm is linear in the size of the image. The simplicity of the algorithm and the low complexity are proved with the interesting improvement of the

ISBN 978-0-9928626-7-1 © EURASIP 2017

compression performance.

V. CONCLUSION

In this paper, a new preprocessing technique for improving the lossless compression of sparse histogram images is presented. The proposed algorithm is very simple and can be attached easily to any lossless image compression codec. This technique exploits the strong correlation between neighboring image blocks to reduce the total image variation and histogram sparseness. The image is divided into blocks and a small set is assigned foreach image block representing its active symbols with an overhead bits as small as possible. Experiments performed for JPEG 2000 show that the lossless compression efficiency is significantly improved. The use of this method for lossless compression of images with higher bit depths can lead to an interesting improvement and can be the subject of a fruital research.

REFERENCES

[1] R. Starosolski, "Compressing images of sparse histograms," in *SPIE Proceedings*, POLAND, 2005, pp. 209–217.

[2] M. Iwahashi, H. Kobayashi, and H. Kiya, "Lossy compression of sparse histogram image," in *Proc of. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, JAPAN, 2012, pp. 1361–1364.

[3] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*, Springer, 2002.

[4] M. W. Marcellin, M. J. Gormish,A. Bilgin, and M. P. Boliek, "An overview of jpeg-2000," in *Proc. IEEE Data Compression Conference*, USA, 2000, pp. 523–541.

[5] M. Boliek, J. S. Houchin, and G. Wu, "Jpeg 2000 next generation image compression system features and syntax," in *Proc. IEEE Int. Conf. on Image Processing*, CANADA, 2000, pp. 45–48.

[6] M. Rabbani and R. Joshi, "An overview of the jpeg2000 still image compression standard," *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 3–48, October 2002.

[7] A. J. Pinho, "An online preprocessing technique for improving the lossless compression of images with sparse histograms," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 5–7, August 2002.

[8] A. Masmoudi and A. Masmoudi, "A new arithmetic coding model for a block-based lossless image compression based on exploiting inter-block correlation," *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1021–1027, July 2015.

[9] M. Aguzzi and M. Albanesi, "A novel approach to sparse histogram image lossless compression using jpeg2000," *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, vol. 5, no. 4, pp. 24–46, July 2006.

[10] M. Aguzzi, "Working with jpeg2000: two proposals for sparse histogram images," in *Proc. IEEE The 12th International Conference on Image Analysis and Processing*, 2003, pp. 408–411.

[11] A. J. Pinho, "On the impact of histogram sparseness on some lossless image compression techniques," in *Proc. of IEEE Int. Conf. on Image Processing(ICIP)*, Greece, 2001, pp. 442–445.

[12] P. J. S. G. Ferreira , and A. J. Pinho, "Histogram packing, total variation, and lossless image compression," in *The 11th European Signal Processing Conference*. IEEE, 2002, pp. 1–4.

[13] P. J. S. G. Ferreira , and A. J. Pinho, "Why does histogram packing improve lossless compression rates?," *IEEE signal processing letters*, vol. 9, no. 8, pp. 259–261, 2002.

[14] A. J. Pinho, "A comparison of methods for improving the lossless compression of images with sparse histograms," in *Proc. of IEEE Int. Conf. on Image Processing(ICIP)*, 2002, pp. 673–676.

[15] A. J. Pinho and A. J. R. Neves, "Block-based histogram packing of color-quantized images," in *Proc of. IEEE Int. Conf. on Multimedia and Expo (ICME)*, USA, 2003, pp. 341–344.

[16] A. Masmoudi, W. Puech, and A. Masmoudi, "An improved lossless image compression based arithmetic coding using mixture of non-parametric distributions," *Multimedia Tools and Applications*, vol. 74, no. 23, pp. 10605–10619, 2015.

[17] A. Masmoudi and A. Masmoudi, "A new arithmetic coding model for a block-based lossless image compression based on exploiting inter-block correlation," *Signal, Image and Video Processing*, vol. 9, no. 5, pp. 1021–1027, 2015.

[18] A. J. Pinho, "Preprocessing techniques for improving the lossless compression of images with quasi-sparse and locally sparse histograms," in *Proc of. IEEE Int. Conf. on Multimedia and Expo (ICME)*, Switzerland, 2002, pp. 633–636.

[19] A. Masmoudi, S. Zouari, and A. Ghribi, "Adaptive block-wise alphabet reduction scheme for lossless compression of images with sparse and locally sparse histograms," *Journal of Electronic Imaging*, vol. 24, pp. 063001, November 2015.

[20] P. J. Ausbeck, "The piecewise-constant image model," *Proceedings of the IEEE*, vol. 88, no. 11, pp. 1779–1789, November 2000.

[21] Universit de Louvain, "An open-source jpeg 2000 codec," http://www.openjpeg.org, April 2014.

[22] B. Carpentieri, M. J.Weinberger, and G. Seroussi, "Lossless compression of continuous-tone images," *Proceedings of the IEEE*, vol. 88, no. 11, pp. 1797–1809, August 2000.

[23] Y.Yoo, Y.G.Kwon, and A.Ortega, "Embedded image-domain adaptive compression of simple images," in *Proc. IEEE The 32nd Asilomar Conf: on Signals, Systems. and Computers*, 1998, vol. 2.

ISBN 978-0-9928626-7-1 © EURASIP 2017