A Lossless Image Coding Method Based on Probability Model Optimization

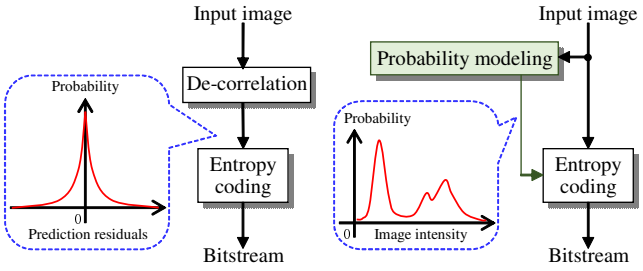
Ichiro Matsuda, Tomokazu Ishikawa, Yusuke Kameda, and Susumu Itoh
Department of Electrical Engineering, Faculty of Science and Technology,
Tokyo University of Science
2641 Yamazaki, Noda-shi, Chiba 278-8510, JAPAN
Email: matsuda@ee.noda.tus.ac.jp

***Abstract*—This paper proposes a novel lossless image coding method which directly estimates a probability distribution of image intensity values on a pel-by-pel basis. In the estimation process, several examples, i.e. a set of pels whose neighborhoods are similar to a local texture of the target pel to be encoded, are gathered from a search window located on an already encoded part of the same image. Then the probability distribution is modeled as a weighted sum of the Gaussian functions whose center positions are given by the individual examples. Furthermore, model parameters that control shapes of the Gaussian functions are numerically optimized so that the resulting coding rate of the image intensity values can be a minimum. Simulation results indicate that the proposed method provides comparable coding performance to the state-of-the-art lossless coding schemes proposed by other researchers.**

***Index Terms*—lossless image coding, template matching, probability model, numerical optimization**

I. Introduction

Most efficient image coding techniques consist of two processing stages: de-correlation and entropy coding. The purpose of the first stage is to remove redundancy of the given image signal by some reversible numerical operations such as linear transform and prediction. In the case of lossless coding, pel-wise adaptive prediction is often employed and no quantization process is applied to the resulting prediction residuals. At the second stage, a probability distribution of the de-correlated signal is estimated by a non-parametric [1] or a parametric [2] probability model, and either Huffman or arithmetic coding is utilized to generate a compressed bitstream under the assumption of the estimated model. In this framework, de-correlated signals obtained through the first stage, namely transform coefficients or prediction residuals, are usually expected to follow a symmetric single-peaked distribution centered around zero as shown in Fig.1 (a). Since the narrower distribution tends to show lower entropy, least squares (LS) [3] or weighted least squares (WLS) [4], [5] is generally used as a criterion for designing the adaptive predictors. By contrast, in our past study, a coding rate of the prediction residuals was iteratively minimized in the predictor design process [6]. In either case, the probability distribution is modeled in a domain of the de-correlated signal and only the single-peaked probability distribution is considered.



(a) Conventional method (b) Proposed method

Fig. 1. Conceptual diagrams of the conventional and proposed methods.

In this paper, we proposed a novel lossless image coding method which integrates the above two processing stages. In other words, the probability distribution is modeled in an image intensity domain without any de-corelation process. A similar idea was presented in [7], where peak positions of the probability distribution were given by multiple linear predictors. This means only local information of the image signal was used to determine the peak positions through the prediction process using causal neighbors. On the other hand, the proposed method exploits non-local information that is gathered from an already encoded area of the same image via a template matching algorithm. Such non-local information can well capture self-similarity inherent in natural images and has been utilized in image denoising [8], [9]. Recently, it is also used in image coding applications [10], [11]. Furthermore, model parameters that control a shape of the probability distribution model are numerically optimized to minimize the resulting coding rate. This framework allows us to model complicated probability distributions having multiple peak positions as shown in Fig.1 (b).

II. Template Matching

In order to exploit non-local information on the image signal, the proposed method performs template matching in a search window placed on the already encoded area. In Fig.2, $\boldsymbol{p}_k \in \mathbb{Z}^2$ indicates a target pel being encoded in the raster scan order and a size of the search window is determined by the maximum distance S from the target pel in horizontal and vertical directions. A template patch is defined by twelve pels

ISBN 978-90-827970-1-5 © EURASIP 2018

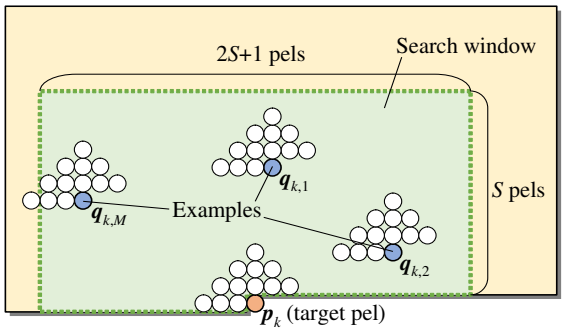


Fig. 2. Template Matching.

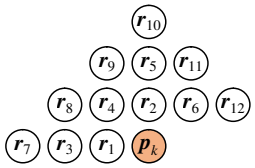


Fig. 3. Arrangement of pels in a template patch.

in a causal neighborhood $\{\boldsymbol{p}_k + \boldsymbol{r}_i \,|\, i = 1, 2, \ldots, 12\}$ with the distance of $||\boldsymbol{r}_i||_1 \leq 3$, where a vector $\boldsymbol{r}_i \in \mathbb{Z}^2$ denotes the position of the i-th pel relative to $\boldsymbol{p}_k$ as shown in Fig.3. For every pel $\boldsymbol{q}$ in the search window, we evaluate similarity of the local textures by the following cost function:

$$J_k(\boldsymbol{q}) = \sum_{i=1}^{12} w_i \cdot \big| f(\boldsymbol{q}+\boldsymbol{r}_i) - \mu(\boldsymbol{q}) - f(\boldsymbol{p}_k+\boldsymbol{r}_i) + \mu(\boldsymbol{p}_k) \big| + \lambda_d \cdot ||\boldsymbol{q} - \boldsymbol{p}_k||_1, \quad (1)$$

where $f(\boldsymbol{q})$ represents an image intensity value at the pel $\boldsymbol{q}$ and $\mu(\boldsymbol{q})$ is a weighted local mean within the adjacent patch:

$$\mu(\boldsymbol{q}) = \sum_{i=1}^{12} w_i \cdot f(\boldsymbol{q} + \boldsymbol{r}_i). \quad (2)$$

In both the equations, a weighing factor w_i is given by the Gaussian function with a constant value of $\sigma_t = 1.25$.

$$w_i = \frac{\exp\left(-\frac{1}{2}||\boldsymbol{r}_i||_1^2/\sigma_t^2\right)}{\sum_{j=1}^{12} \exp\left(-\frac{1}{2}||\boldsymbol{r}_j||_1^2/\sigma_t^2\right)}. \quad (3)$$

The first term of the right hand side of Eq.(1) can be seen as a weighted version of the Zero-mean Sum of Absolute Differences (ZSAD) [11]. We use Eq.(3) to give heavier weights around the target pel and its counterpart. Moreover, the second term is added with the intention of giving small priority to spatially near positions in the search window.

As a result of this template matching, M pels are gathered in ascending order of the cost function and considered to be a set of examples containing beneficial non-local information: $\boldsymbol{E}_k = \{\boldsymbol{q}_{k,1}, \boldsymbol{q}_{k,2}, \ldots, \boldsymbol{q}_{k,M}\}$. From these examples, M estimations of the image intensity values are calculated by compensating their local means:

$$f_{k,m} = f(\boldsymbol{q}_{k,m}) - \mu(\boldsymbol{q}_{k,m}) + \mu(\boldsymbol{p}_k), \ (m = 1, 2, \ldots, M). \quad (4)$$

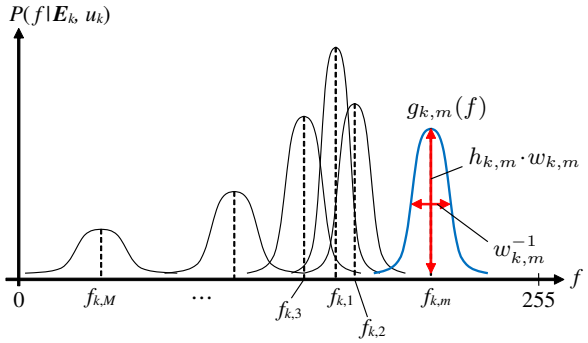


Fig. 4. Probability distribution model.

III. Probability Distribution Model

The above estimation $f_{k,m}$ obtained from the past example $\boldsymbol{q}_{k,m}$ is expected to be close to $f(\boldsymbol{p}_k)$ if the corresponding cost function $d_{k,m} = J_k(\boldsymbol{q}_{k,m})$ is sufficiently small. According to this expectation, a probability distribution of the image intensity f at the target pel $\boldsymbol{p}_k$ under the condition of given $\boldsymbol{E}_k$ is modeled as weighted sum of the Gaussian functions.

$$\Pr(f \,|\, \boldsymbol{E}_k, u_k) \propto P(f \,|\, \boldsymbol{E}_k, u_k) = \sum_{m=1}^{M} g_{k,m}(f) + \varepsilon, \quad (5)$$

$$g_{k,m}(f) = h_{k,m} \cdot w_{k,m} \cdot \exp\left(-w_{k,m}^2 \cdot (f - f_{k,m})^2\right), \quad (6)$$

where u_k is a locally calculated feature quantity related to context of the target pel $\boldsymbol{p}_k$ and its definition will be described later. In addition, ε is a small positive constant to avoid zero probabilities and set to 2^{-20} in this paper. The shape of the Gaussian function $g_{k,m}(f)$ is determined by three parameters $h_{k,m}$, $w_{k,m}$ and $f_{k,m}$ which respectively associated with height, width and a peak position of the Gaussian function as shown in Fig.4. We consider that the parameter $h_{k,m}$ has a connection with reliability of the example $\boldsymbol{q}_{k,m}$, and therefore define it as a parametric function of $d_{k,m}$.

$$h_{k,m} = \exp(-a_1 \cdot d_{k,m}). \quad (7)$$

Similarly, the parameter $w_{k,m}$ is considered to be affected by u_k in addition to $d_{k,m}$, and defined as:

$$w_{k,m} = a_0 \cdot \exp(-a_2 \cdot d_{k,m}) \cdot \exp(-a_3 \cdot u_k). \quad (8)$$

Consequently, occurrence probabilities for all the possible values of $f(\boldsymbol{p}_k)$ can be calculated by normalizing Eq.(5). In the case of an 8-bit grayscale image, we obtain

$$\Pr(f(\boldsymbol{p}_k) \,|\, \boldsymbol{E}_k, u_k) = \frac{P(f(\boldsymbol{p}_k) \,|\, \boldsymbol{E}_k, u_k)}{\sum_{f=0}^{255} P(f \,|\, \boldsymbol{E}_k, u_k)}. \quad (9)$$

ISBN 978-90-827970-1-5

By assuming these probabilities, the number of coding bits required for entropy coding of the actual value of $f(\boldsymbol{p}_k)$ can be approximated as:

$$L(\boldsymbol{p}_k) = -\log_2 \Pr(f(\boldsymbol{p}_k) \,|\, \boldsymbol{E}_k, u_k) = \frac{1}{\ln 2}\left[\ln\left(\sum_{f=0}^{255} P(f|\boldsymbol{E}_k, u_k)\right) - \ln P(f(\boldsymbol{p}_k)|\boldsymbol{E}_k, u_k)\right]. \tag{10}$$

This quantity reflects local fitness of the given model and is utilized for the definition of u_k:

$$u_k = \sum_{i=1}^{12} w_i \cdot L(\boldsymbol{p}_k + \boldsymbol{r}_i). \tag{11}$$

IV. Optimization of Model Parameters

The shape of the above probability distribution model is controllable by four parameters $\{a_n \,|\, n = 0, \ldots, 3\}$. Here, we want to optimize these model parameters to reduce the resulting coding rate in a certain region $\boldsymbol{\Omega}$. Hence, we define the objective function as:

$$J(\boldsymbol{\Omega}) = \sum_{\boldsymbol{p}_k \in \boldsymbol{\Omega}} L(\boldsymbol{p}_k) + \lambda_p \sum_{n=0}^{3} (a_n - \alpha_n)^2, \tag{12}$$

where α_n is an initial value of the model parameter a_n and λ_p is a weight of the regularizer which prevents deviation of the parameters from their initial values. We use a fixed value of $\lambda_p = 0.1$ in this paper. Minimization of the above objective function is formulated as a nonlinear optimization problem with respect to the four model parameters. The objective function is differentiable by each model parameter and its gradient component can be calculated by:

$$\frac{\partial}{\partial a_n} J(\boldsymbol{\Omega}) = \frac{1}{\ln 2} \sum_{\boldsymbol{p}_k \in \boldsymbol{\Omega}} \left(\frac{\sum_{f=0}^{255} \frac{\partial}{\partial a_n} P(f|\boldsymbol{E}_k, u_k)}{\sum_{f=0}^{255} P(f|\boldsymbol{E}_k, u_k)} - \frac{\frac{\partial}{\partial a_n} P(f(\boldsymbol{p}_k)|\boldsymbol{E}_k, u_k)}{P(f(\boldsymbol{p}_k)|\boldsymbol{E}_k, u_k)} \right) + 2\lambda_p (a_n - \alpha_n), \tag{13}$$

where

$$\frac{\partial}{\partial a_0} P(f \,|\, \boldsymbol{E}_k, u_k) = \frac{1}{a_0} \sum_{\boldsymbol{p}_k \in \boldsymbol{\Omega}} \left[1 - 2(f - f_{k,m})^2 \cdot w_{k,m}^2\right] \cdot g_{k,m}(f), \tag{14}$$

in the case of $n = 0$, for example. Therefore, we can use a gradient-based numerical optimization algorithm such as the quasi-Newton method [12]. In practice, we divide the image into square blocks of $N \times N$ pels, and optimize the model parameters in each block by regarding the block as the region $\boldsymbol{\Omega}$. Strictly speaking, the feature quantity u_k also depends on the model parameters and it must be considered in calculation of Eq.(14). For simplicity, however, the quasi-Newton method is performed under the condition of fixed u_k and alternately iterated with re-calculation of Eq.(11) for all the pels in the block $\boldsymbol{\Omega}$. Finally, the optimized model parameters are linearly quantized with 8-bit accuracy in the ranges shown in Table I, and encoded as side-information.

TABLE I
Quantization ranges and initial values of the model parameters.

Parameter	Quantization range	Initial value (α_n)
a_0	[0.0, 8.0]	2.0
a_1	[−1.0, 3.0]	1.0
a_2	[−1.0, 2.0]	0.5
a_3	[−1.0, 2.0]	0.5

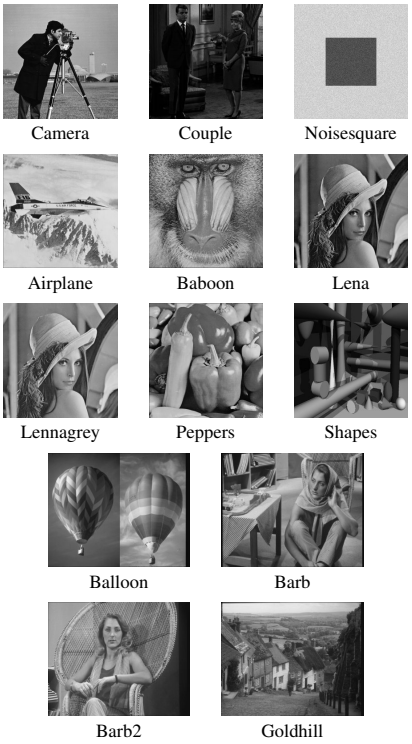


Fig. 5. Test images.

V. Experimental Results

The proposed method was implemented with a fast multi-symbol arithmetic coding technique called range coder [13] and tested for grayscale images shown in Fig.5[1].

In the template matching, the search window size S and the weight λ_d for the distance term in Eq.(1) are important parameters affecting properties of the obtained examples. We therefore investigated proper combinations of both parameters in terms of the average coding rates. In this experiment, the side length of the square block $\boldsymbol{\Omega}$ and the number of the examples are fixed to $N = 64$ and $M = 64$, respectively. As we can see in Fig.6, better results are obtained when $S > 72$ and $\lambda_d > 0.01$. According to this observation, the settings of $S = 80$ and $\lambda_d = 0.03$ will be adopted in the rest of the paper.

The number of examples $M = |\boldsymbol{E}_k|$ is another important parameter concerning accuracy of the probability model. Figure 7 plots relationship between the average coding rate and the parameter M when the side length of the

[1]This dataset was used in [7] and is currently available via Internet Archive: https://web.archive.org/web/*/http://www.csse.monash.edu.au/~bmeyer/tmw/

ISBN 978-90-827970-1-5

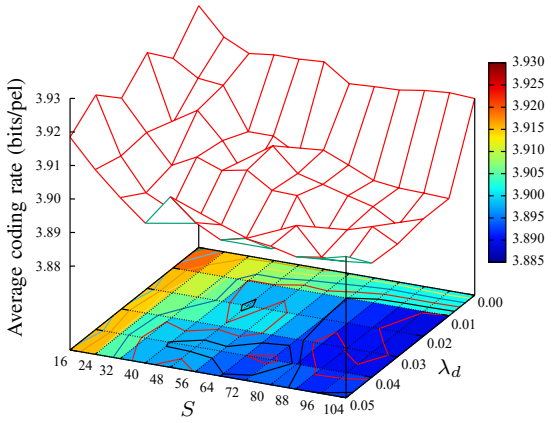


Fig. 6. Average coding rate as a function of two parameters S and λ_d.

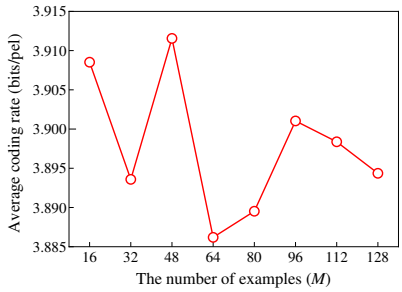


Fig. 7. Average coding rate as a function of the number of examples.

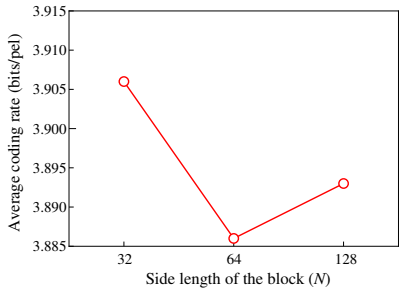
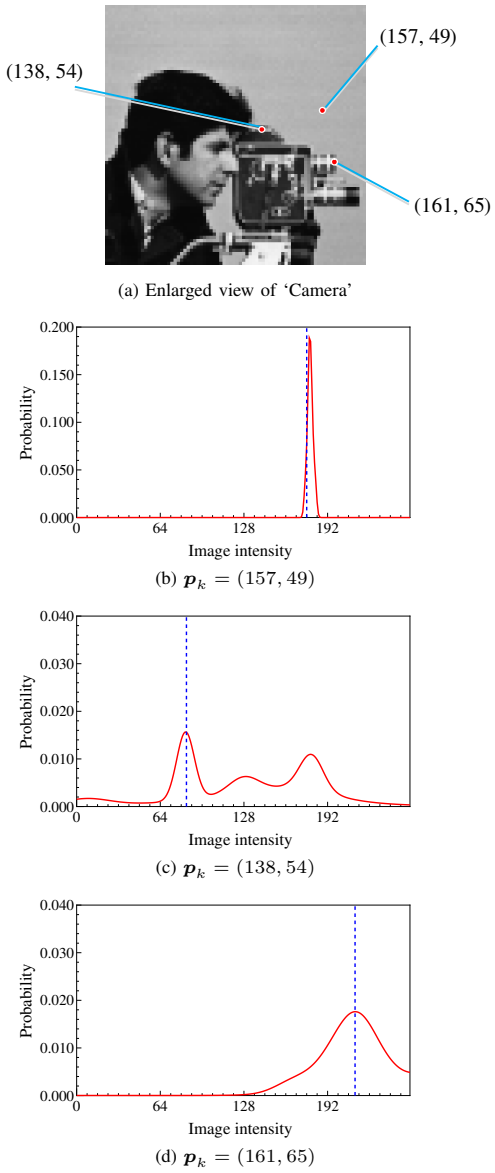


Fig. 8. Average coding rate as a function of the block size.



(a) Enlarged view of 'Camera'

(b) $\boldsymbol{p}_k = (157, 49)$

(c) $\boldsymbol{p}_k = (138, 54)$

(d) $\boldsymbol{p}_k = (161, 65)$

Fig. 9. Examples of the estimated probability distributions.

block $\boldsymbol{\Omega}$ is set to $N = 64$. The coding efficiency is rather sensitive to this parameter mainly due to unstable behavior of the optimization process for unnatural images such as 'Noisesquare'. Nevertheless, $M = 64$ seems to be a reasonable choice in the average sense.

Furthermore, we tested different size of the block $\boldsymbol{\Omega}$ while setting $M = 64$. It is shown in Fig.8 that optimizing the model parameters for each block of 64×64 pels gives better trade-off between the number of coding bits spent for the image signal and the amount of the side-information on the model parameters.

Figure 9 shows some examples of the probability distributions estimated by the proposed method. A blue dotted line indicates an actual image intensity value at each sample position. We can see that single-peaked distributions with different sharpness as well as a much complicated distribution with multiple peaks can be modeled by the proposed method.

Finally, coding rates of the proposed method are compared with the other lossless coding schemes: MRP (version 0.5) [6], Vanilc WLS D (version 1.0) [14], TMW (version 0.51) [7], Glicbawls [5], WebP lossless (version 0.6.0) [15], FLIF (version 0.3, non-interlaced) [16], JPEG-LS [17] and JPEG 2000 [18]. In Table II, the best and the second best coding rates are shown as **bold** and underlined numbers for each image, respectively. It is demonstrated that the proposed method attains comparable coding performance to Vanilc and TMW which are known as the state-of-the-art lossless image coding schemes. However, there is a certain performance gap against our MRP method which had been drastically improved from its earlier version [19].

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE II
COMPARISON OF CODING RATES (BITS/PEL).

Image	Size	Proposed	MRP	Vanilc WLS D	TMW	Glicbawls	WebP lossless	FLIF	JPEG-LS	JPEG 2000
Camera	256×256	3.960	**3.949**	3.995	4.098	4.208	4.274	4.285	4.314	4.535
Couple		3.415	**3.388**	3.459	3.446	3.543	3.703	3.677	3.699	3.915
Noisesquare		5.298	5.270	**5.159**	5.542	5.415	5.203	5.335	5.683	5.634
Airplane	512×512	3.632	3.591	**3.575**	3.601	3.668	3.894	3.794	3.817	4.013
Baboon		5.727	**5.663**	5.678	5.738	5.666	5.891	6.078	6.037	6.107
Lena		4.330	4.280	**4.246**	4.300	4.295	4.514	4.642	4.607	4.684
Lennagrey		3.944	3.889	**3.856**	3.908	3.901	4.145	4.252	4.238	4.303
Peppers		4.267	4.199	**4.187**	4.251	4.246	4.495	4.595	4.513	4.629
Shapes		0.715	**0.685**	1.302	0.740	2.291	1.023	0.722	1.214	1.926
Balloon	720×576	2.673	**2.579**	2.626	2.649	2.640	2.925	2.856	2.904	3.031
Barb		3.997	**3.815**	**3.815**	4.084	3.916	4.547	4.500	4.691	4.600
Barb2		4.287	**4.216**	4.231	4.378	4.318	4.668	4.656	4.686	4.789
Goldhill		4.276	**4.207**	4.229	4.266	4.276	4.464	4.518	4.477	4.603
Average		*3.886*	***3.826***	*3.874*	*3.923*	*4.030*	*4.134*	*4.147*	*4.222*	*4.367*

VI. CONCLUSION

In this paper, we have proposed a lossless image coding method based on optimization of the probability model used for entropy coding of image intensity values. The model is defined by the weighted sum of the Gaussian functions whose peak positons are determined via template matching on the already encoded causal area. Moreover, several model parameters that control height and width of the individual Gaussian functions are numerically optimized to minimize the resulting coding rate on a block-by-block basis. Experimental implementation of the proposed method provided promising coding performance, though its framework is quite different from the conventional schemes typically based on prediction techniques. It is worth noting that some of the examples used in the probability modeling can be replaced by predicted values calculated from local neighbors [20]. Collaboration with efficient prediction techniques such as [6] will be a part of our future studies.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 17K00247.

REFERENCES

[1] 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, Dec. 2015.

[2] H. Ye, G. Deng and J. C. Devlin, "Parametric Probability Models for Lossless Coding of Natural Images," *Proc. of 11th European Signal Processing Conference (EUSIPCO 2002)*, Vol.II, pp.514–517, Sep. 2002.

[3] L.-J. Kau and Y.-P. Lin, "Adaptive Lossless Image Coding Using Least Squares Optimization With Edge-Look-Ahead ," *IEEE Trans. on Circuits and Systems II: Express Briefs*, Vol.52, No.11, pp.751–755, Nov. 2005.

[4] H. Ye, G. Deng and J. C. Devlin, "A Weighted Least Squares Method for Adaptive Prediction in Lossless Image Compression," *Proc. of Picture Coding Symposium (PCS 2003)*, pp.489–493, Apr. 2003.

[5] B. Meyer and P. Tischer, "Glicbawls – Grey Level Image Compression by Adaptive Weighted Least Squares," *Proc. 2001 Data Compression Conference (DCC 2001)*, p.503, Mar. 2001.

[6] I. Matsuda, N. Ozaki, Y. Umezu and S. Itoh, "Lossless Coding Using Variable Blok-Size Adaptive Prediction Optimized for Each Image," *Proc. of 13th European Signal Processing Conference (EUSIPCO 2005)*, Sep. 2005.

[7] B. Meyer and P. Tischer, "TMW – A New Method for Lossless Image Compression", *Proc. of 1997 Picture Coding Symposium (PCS '97)*, pp.533–538, Sep. 1997.

[8] A. Buades, B. Coll and J.-M. Morel, "A Non-Local Algorithm for Image Denoising," *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, pp.60–65, June 2005.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Trans. on Image Prpcessing*, Vol.16, No.8, pp.2080–2095, Aug. 2007.

[10] E. Wige, G. Yammine, P. Amon, A. Hutter and A. Kaup, "Pixel-Based Averaging Predictor for HEVC Lossless Coding", *IEEE International Conference on Image Processing (ICIP 2013)*, pp.1806–1810, Sep. 2013.

[11] N. Nakajima, Y. Kameda, I. Matsuda and S. Itoh, "Example-Based Context-Adaptive Probability Modeling for Lossless Image Coding," *Proc. of International Workshop on Smart Info-Media Systems in Asia (SISA 2014)*, SS2-02, pp.95–98, Oct. 2014.

[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B.P. Flannery, "Numerical Recipes: The Art of Scientific Computing, Third Edition," *Cambridge University Press*, 2007.

[13] M. Schindler, "A Fast Renormalization for Arithmetic Coding," *Proc. of Data Compression Conference (DCC '98)*, p.572, 1998.

[14] A. Weinlich, P. Amon, A. Hutter and A. Kaup, "Probability Distribution Estimation for Autoregressive Pixel-Predictive Image Coding," *IEEE Trans. on Image Processing*, Vol.25, No.3, pp.1382–1395, Mar. 2016.

[15] Google Developers, "A New Image Format for the Web," http://developers.google.com/speed/webp/.

[16] J. Sneyers and P. Wuille, "FLIF: Free Lossless Image Format Based on MANIAC Compression ," *Proc. of IEEE International Conf. on Image Processing (ICIP 2016)*, pp.66–70, Sep. 2016.

[17] ISO/IEC, 14495-1:1999, "Information Technology — Lossless and Near-lossless Compression of Continuous-tone Still Images: Baseline," Dec. 1999.

[18] ITU-T Rec. T.800 | ISO/IEC 15444-1, "Information Technology — JPEG 2000 Image Coding System — Part 1: Core Coding System," 2001.

[19] I. Matsuda, H. Mori and S. Itoh, "Design of a Minimum-Rate Predictor and its Application to Lossless Image Coding," *Proc. of 10th European Signal Processing Conference (EUSIPCO 2000)*, Vol.II, pp.1205–1208, Sep. 2000.

[20] T. Sumi, Y. Inamura, Y. Kameda, T. Ishikawa, I. Matsuda and S. Itoh, "Lossless Image Coding Based on Probability Modeling Using Template Matching and Linear Prediction," *IEICE Trans. on Fundamentals*, Vol.E100-A, No.11, pp.2351–2354, Nov. 2017.

ISBN 978-90-827970-1-5 © EURASIP 2018