DESIGN AND IMPLEMENTATION OF NON-UNIFORM QUANTIZERS FOR DISCRETE INPUT SAMPLES AND ITS APPLICATION TO AN IMAGE PROCESSING ALGORITHM

Nikola Simić[1], Zoran H. Perić[1], Milan Savić[2]

[1]University of Niš, Faculty of Electronic Engineering,
Department of Telecommunications, Niš, Republic of Serbia
[2]University of Pristina, Faculty of Natural Science and Mathematics,
Department of Informatics, Kosovska Mitrovica, Republic of Serbia

Abstract. *This paper describes an algorithm for grayscale image compression based on non-uniform quantizers designed for discrete input samples. Non-uniform quantization is performed in two steps for unit variance, whereas design is done by introducing a discrete variance. The best theoretical and experimental results are obtained for those discrete values of variance which provide the operating range of quantizer located in the vicinity of maximal signal value that can appear on the entrance. The experiment is performed by applying proposed quantizers for compression of standard test grayscale images as a classic example of discrete input source. The proposed fixed non-uniform quantizers, designed for discrete input samples, provide up to 4.93* [dB] *higher* PSQNR *compared to the fixed piecewise uniform quantizers designed for discrete input samples.*

Key words: *Discrete input samples, grayscale image processing, non-uniform quantization, optimal input range.*

1. INTRODUCTION

The interest in methods of digital image processing comes from two basic ideas. First of all, rapidly growing information systems aim at reducing the amount of data required for data processing in order to use narrower bandwidth, as well as to save available storage. Next, visual interpretation has to be improved since digital images are widely used in a number of applications [1]. Generally, all compression algorithms may be classified in two groups – 'lossless' compression algorithms if there is no loss of information, and 'lossy' methods if some information is lost irreversibly [1], [2]. Even though there is a variety of compression algorithms for different purposes [3], research areas are still expanding. In recent years, schemes which incorporate compressive sensing became very important and

some restoration as well as image reconstruction schemes made an impact in the image processing field [4]-[5]. Besides schemes developed for software applications, some effort is also paid to FPGA based solutions [6]. This paper deals with a type of improved BTC (Block Truncation Coding) algorithm that is a kind of a 'lossy' method, used for compression of grayscale images [7]. Although the basic algorithm has been well-known for years, some upgrades proposed in recent years have found application in modern systems [8]. Moreover, an improved block truncation coding algorithm based on optimized dot diffusion was proposed by Guo *et*. al [9], whereas an effective image retrieval system was presented a year later [10]. Also, a data hiding scheme based on BTC algorithm, designed to embed a huge amount of watermarks was presented in the paper [11], so it can be concluded that the core algorithm can be still improved and implemented in modern systems. Despite the core algorithm and its modifications usually can not improve the coding gain comparing to the modern state-of-the-art techniques such as jpeg and jpeg2000, the computational complexity of those schemes is much lower compared to the aforementioned state-of-the-art solutions, which makes it very suitable for image retrieval purposes [10].

The difference in designing of fixed uniform quantizers for continual and discrete input was observed in papers [12], [13]. Further research in this direction included designing of fixed piecewise uniform quantizers described in [14]. This paper is a logical continuation of the research. We expect that the gain due to different non-uniform quantizer designing for discrete and continual input is higher than the maximal difference of *PSQNR* (Peak Signal-to-Quantization-Noise Ratio) between fixed piecewise uniform (L=16) and optimal non-uniform quantizer that is equal to 0.7 [dB] (for continual input signal)[14], [15]. The proposed design is fixed, it was tested for a set of standard test grayscale images and optimal parameters are found. However, non-uniform quantizer can be designed by using Lloyd-Max algorithm which represents a very powerful iterative solution [16]. Moreover, high-quality performance can be achieved by introducing variance adaptation which would provide better quality of reconstructed image [17]. On the other hand, the proposed design is less complex and it requires less processing time, as it represents a kind of fixed scalar quantization.

The paper is organized as follows. In Section 2 basic modelling of discrete input source is shown, improved by introducing non-uniform quantization. Section 3 describes an algorithm for grayscale image compression that is used for experimental analysis. Finally, the obtained theoretical and experimental results as well as the obtained gain in comparison to other models are presented in Section 4.

2. SYSTEM MODEL

The considered system consists of two stages − the purpose of uniform quantizer Q_0 exploited in the first stage is to convert analog input signal to discrete samples, whereas the proposed quantizer Q, designed for discrete input, is exploited in the second stage in order to perform additional data compression.

In the first step, samples with a continual amplitude have to be quantized with a fixed uniform quantizer Q_0 which is described with N_0 output levels, $X=\{x_1,x_2,\ldots,x_{N_0}\}$, and the maximal amplitude $x_{\max}$, which depends on the input signal range [14], [17]. Considered pixel values of standard grayscale images are described with 8 bits and they can take values from 0 to 255, so $x_{N_0} = 255$. Furthermore, quantization process in BTC algorithm is

based on quantization of distinction between the original and mean pixel value of all pixels in a block. Therefore, the number of output levels N_0 is equal to 512. On the other hand, samples with continuous amplitude can be described only as random variables, since the input information is unknown. In probability theory, random variables are described by using probability density function (PDF) which provides the relative likelihood for the observed random variable to take on a given value. So far, it is shown in literature that Laplacian source ensures good matching between a BTC model and reality [1], [7]. Consequently, in the rest of the paper we will suppose that the information source is Laplacian with a memoryless property and mean value equal to zero. It is defined with:

$$p(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}\,|x|}{\sigma}\right),$$ (1)

where σ represents a standard deviation of the random variable x.

The second step of quantization process involves quantization of discrete output samples from the quantizer Q_0 using N quantization levels, where $N < N_0$. Probabilities of these discrete input levels for Laplacian distribution are:

$$P_i = \int_{x_i}^{x_{i+1}} p(x)\mathrm{d}x = \frac{1}{2}\left(\exp\left(-\frac{\sqrt{2}x_i}{\sigma}\right) - \exp\left(-\frac{\sqrt{2}x_{i+1}}{\sigma}\right)\right),$$ (2)

where $i = 0, \dots, N_0 - 1$.

The main goal of this phase is additional data compression. In the rest of the paper the quantizer from the second step is denoted with Q. This paper deals with designing and optimization of quantizer Q. So far in literature, the application of both uniform and piecewise uniform quantizers was described, and in this paper we propose application of a non-uniform quantizer since it provides better quality of reconstructed signal for the equal number of quantization levels [1].

The design of the non-uniform quantizer Q is done as follows. Firstly, we design the optimal compandor with N quantization levels for the unit standard deviation ($\sigma = 1$). Its compressor function maps the range $(-\infty, \infty)$ to $(-1, 1)$. The compressor function formed in this way can be defined with:

$$c(x) = -1 + 2\frac{\int_{-\infty}^{x} p^{1/3}(t)\mathrm{d}t}{\int_{-\infty}^{\infty} p^{1/3}(t)\mathrm{d}t}.$$ (3)

Decision thresholds obtained in this way can be calculated as [15]:

$$t_i = \frac{3}{\sqrt{2}}\log\left(\frac{2i}{N}\right) \qquad 0 \le i \le \frac{N}{2},$$ (4)

$$t_i = \frac{3}{\sqrt{2}}\log\left(\frac{N}{2(N-i)}\right) \qquad \frac{N}{2} < i \le N.$$ (5)

Furthermore, representational levels are determined with [15]:

$$\omega_i = \frac{3}{\sqrt{2}} \log\left(\frac{2i}{N} - 1\right) \qquad 1 \le i \le \frac{N}{2}, \tag{6}$$

$$\omega_i = \frac{3}{\sqrt{2}} \log\left(\frac{N}{2(N-1)+1}\right) \qquad \frac{N}{2} \le i \le N. \tag{7}$$

In all previous equations $\log(x)$ represents natural logarithm of x. The range of quantizer designed in this way is $(-t_N, t_N)$. Since the obtained range is not adjusted to the theoretical range of pixel values, denormalization is required.

Due to the fact that for a low number of quantization levels $t_N < x_{N_0}$ [15] is always valid, denormalization is performed by introducing a discrete variance $\hat{\sigma}_d$. It is obtained by multiplying decision thresholds t_i and representational levels ω_i with discrete variance $\hat{\sigma}_d$ that is used for quantizer designing. Finally, decision thresholds and representational levels of quantizer Q are determined with:

$$x_i = t_i \cdot \hat{\sigma}_d, \quad 0 \le i < N, \tag{8}$$

$$y_i = \omega_i \cdot \hat{\sigma}_d, \quad 1 \le i \le N. \tag{9}$$

The maximal support x_N can be defined in different ways [15], and in this paper we have decided to choose a simplest form in order to place the last represent at the half of the decision range. However, in the case if $x_N < x_{N_0}$, the overload distortion will exist. On the other hand, if $x_N > x_{N_0}$, the range $[x_{N_0}, x_N]$ will be unused. As a result, higher granular distortion would exist.

If the system conditions require designing of fixed quantizer with the unused range (case $x_N > x_{N_0}$), we propose additional modification by introducing another denormalization parameter Δ. Its function is to adapt the range $[-x_N, x_N]$ formed in the previous step, to the range $[-x_r, x_r]$, where the desired maximal value of the range is denoted with x_r. Consequently, we define parameter Δ with:

$$\Delta = x_r / x_N. \tag{10}$$

Finally, decision thresholds and representational levels of quantizer Q in the case $x_N > x_{N_0}$ are equal to:

$$x_i' = x_i \cdot \Delta, \quad 0 \le i \le N, \tag{11}$$

$$y_i' = y_i \cdot \Delta, \quad 1 \le i \le N. \tag{12}$$

As this is a kind of a 'lossy' compression method, some information will be lost irreversibly during the quantization process. As a standard measure of a reconstructed signal quality we estimate distortion (D) which consists of both granular (D_g) and overload (D_o) distortion that can be calculated with [8], [9]:

$$D_g = 2 \sum_{i=1}^{N/2} \sum_{j=1}^{k_i} (x_{ij} - y_i)^2 P(x_{ij}), \tag{13}$$

$$D_0 = 2 \sum_{j=1}^{s} (x_j - y_{N/2})^2 P(x_j). \tag{14}$$

In Eq. (13) parameter k_i denotes the number or input levels mapped with y_i whereas $x_{ij} \in X$. Moreover, in Eq. (14) $x_j \in X$, parameter s denotes the total number of pixel values from the theoretical range, which are not placed within the designed range. This parameter can be calculated as:

$$s = x_{N_0} - x_N. \tag{15}$$

Finally, the total distortion is equal to:

$$D_t = D_g + D_o. \tag{16}$$

3. ALGORITHM FOR IMAGE PROCESSING

The proposed design of second-stage quantizer Q from Section 2 is tested by analyzing its application to the image processing algorithm, defined as follows.

1. The image is divided into M non-overlapping blocks of dimensions $m \times m$.

2. Each block is processed separately by sending data and reconstructing information at the receiver side. The algorithm processes pixels from left to right and from top to bottom.

3. The mean value of all pixels in the block (x_{av}) is calculated and then quantized ($\hat{x}_{av}$) with a fixed uniform quantizer. In order to minimize the error in the reconstruction process, coding process uses values which are available to the decoder.

4. The difference blocks of $m \times m$ pixels are formed. Elements of a block are denoted with $d_{i,j}$ and obtained as:

$$d_{i,j} = x_{i,j} - \hat{x}_{av}, \tag{17}$$

where $x_{i,j}$ is original pixel value and $i = 1,\ldots, m$; $j = 1,\ldots, m$. Elements of a difference block have Laplacian distribution [1], and they can take integer values $[-x_{N_0}, x_{N_0}]$.

5. Elements of difference blocks are quantized by using proposed fixed non-uniform quantizers from Section 2. These elements are denoted with $\hat{d}_{i,j}$, coded with $\log(N)$ bits and transmitted to the receiver.

6. In the receiver, pixel reconstruction is done as:

$$\hat{x}_{i,j} = \hat{d}_{i,j} + \hat{x}_{av}. \tag{18}$$

During quantization process there was made distortion of original image in step 5. It can be experimentally measured as [9]:

$$D = \frac{1}{m \times m} \sum_{i=1}^{m} \sum_{j=1}^{m} (x_{i,j} - \hat{x}_{i,j})^2 = \frac{1}{m \times m} \sum_{i=1}^{m} \sum_{j=1}^{m} (d_{i,j} - \hat{d}_{i,j})^2. \tag{19}$$

The flow chart of this algorithm is shown in Fig. 1.
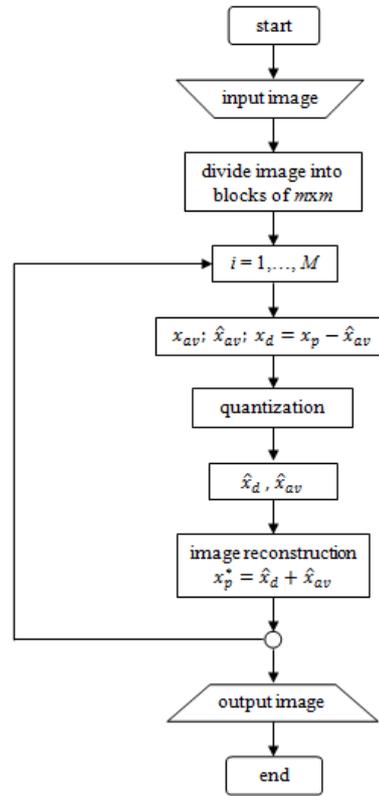
Fig. 1 Flow chart of the proposed grayscale image compression method

4. NUMERICAL RESULTS

To demonstrate the performance of the proposed algorithm for image compression, we will show a comparison of theoretical with experimental results obtained for a set of standard test grayscale images as well as a comparison with the results available in literature for piecewise uniform quantization model [14]. All theoretical calculations and experimental results are done for a set of three standard test grayscale images (Lena, Street and Boat).

We estimate system performance using average bit-rate R_b and *PSQNR* which represent standard measures. Since we discuss fixed non-uniform quantizers, average bit-rate depends on the number of quantization levels N and the number of bits required for transmitting the mean value $\hat{x}_{av}$. On the other hand, *PSQNR* is defined with [13], [14], [17]:

$$PSQNR = 10\log_{10}\left(\frac{x_{N_0}^2}{D}\right) \text{ [dB]} \tag{20}$$

For measuring experimental $PSQNR_{ex}$ we use Eq.(20), whereas D is defined with Eq. (19). However, theoretical results have to include weighting function, since input samples do not occur with the same probabilities [14]. The weighting function in linear domain for tested images is shown in Fig. 2.
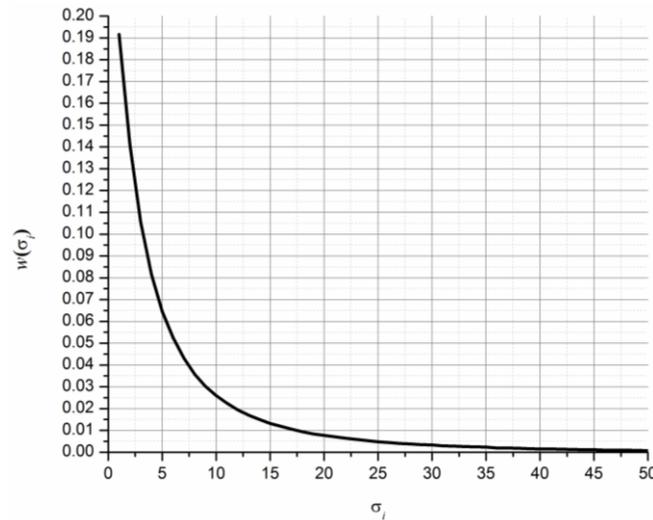
Fig. 2 The weighting function

In Fig. 2, σ_i represents standard deviation of the difference between pixels and the mean value of the block that pixel belongs to.

Taking previous consideration into account, including weighting averaging for the observed test grayscale images and considering that total distortion is defined with Eq.(16), theoretical results are denoted with $PSQNR_{wav}$. This measure is defined with [14]:

$$PSQNR_{wav} = \sum_{\sigma_i=1}^{255} w(\sigma_i)PSQNR(\sigma_i) \text{ [dB]} \tag{21}$$

Table 1 shows obtained experimental results of applying the proposed algorithm for grayscale image compression as well as corresponding theoretical results. It can be seen that experimental results very well follow changes of theoretical values, whereas relative difference between theoretical and experimental values occurs due to non-ideal modelling with Laplacian source as well as because of averaging for a set of images [18]. From Table 1, it can be clearly seen that the best theoretical and experimental results are obtained for those values of discrete variances ($\hat{\sigma}_d = 17$ for $N = 32$ and $\hat{\sigma}_d = 15$ for $N = 64$) which ensures input range of quantizer Q as close as possible to the range $(-152, 152)$ [14], [17]. Consequently, this means that parameter $x_r = 152$.

Table 1 Comparison of experimental and theoretical results for the proposed model

N	$\hat{\sigma}_d$	$PSQNR_{wav}$[dB]	$PSQNR_{ex}$[dB]	x_N	R_b [bpp]
	15	46.82	47.57	132	
32	17	46.43	46.94	149	5.375
	29	44.59	44.51	255	
	15	49.38	51.57	154	
64	24	49.00	50.85	247	6.375
	29	48.01	48.50	298	

Moreover, it can be noticed that for the case $N = 64$ and $\hat{\sigma}_d = 29$, overload distortion does not exists since the range (-298, 298) is wider of the theoretical range (-255, 255) and the support region is not adapted to the theoretical one. In this case, decision thresholds and representational levels could be calculated using Eqs.(10)-(12). However, this modification involves additional hardware requirements and processing time as well as information about x_r for specific systems regarded to the nature of the input signal. In Fig. 3 we have shown original test grayscale images of resolution 512×512 pixels, while in Fig. 4 we have presented corresponding images from Fig. 3, after processing with the proposed algorithm for N=32 quantization levels and $\hat{\sigma}_d = 15$.

(a) (b) (c)
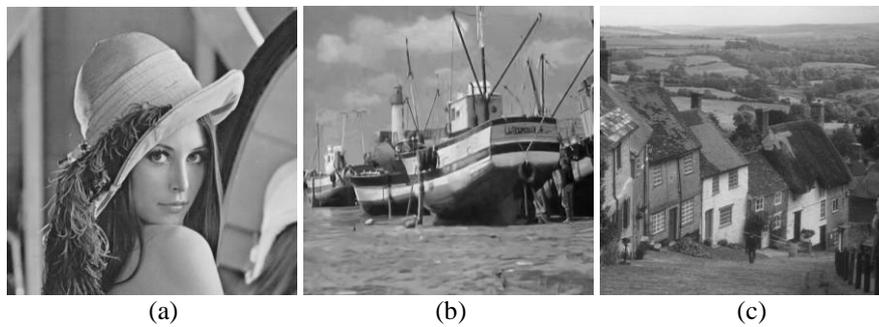
Fig. 3 Standard test grayscale images: (a) Lena, (b) Boat and (c) Street

(a) (b) (c)
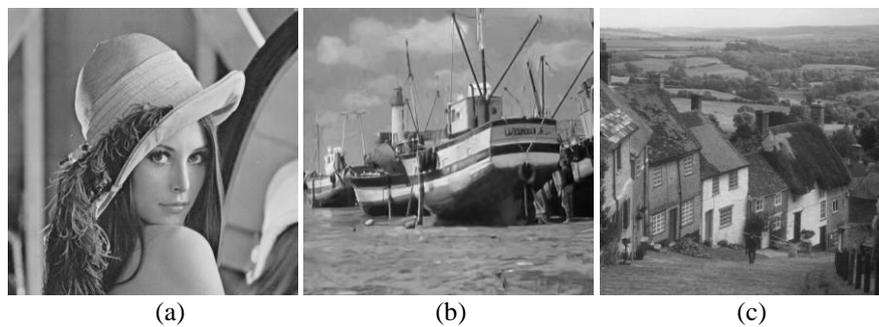
Fig. 4 Standard test grayscale images from Fig. 3, after compression
with the proposed algorithm (N=32): (a) Lena, (b) Boat and (c) Street

In order to compare the obtained results with models available in the literature, we perform comparison of both experimental and theoretical results with system performance of the model based on fixed piecewise uniform quantizers designed for discrete input, as it represents the model with similar complexity. The experimental comparison is measured as experimental gain of the proposed method and it represents the difference of *PSQNR* between the proposed and equivalent results from Savic et al. [14], i.e. Gain [dB] = $PSQNR_{ex}$[dB] - $PSQNR_{eq(N)}^{inf}$ [dB], where equivalent results are provided for *N*=32 and *N*=64 quantization levels. In [14], obtained experimental results as close as to the non-uniform quantization are achieved for $N = 32$ and $L = 16$ ($PSQNR_{ex(32)}^{inf} = 42.64$ [dB], $R_b = 5.375$ [bpp]), whereas corresponding theoretical performance is $PSQNR_{th(32)}^{inf}$=42.29 [dB]. Since the paper [14] did not deal with systems that use $N = 64$ levels, comparison for these results is done considering the rule that *PSQNR* values increase/decrease for 5.5 [dB] by changing the bit-rate for 1 bit [13], [14]. Respecting that bit-rate difference between quantizers that are designed for $N = 32$ and $N = 64$ quantization levels is 1 [bpp], corresponding result for $N = 64$, which is used for comparison, is $PSQNR_{eq(64)}^{inf} =$ 42.640+1*5.5 = 48.14 [dB]. Comparing the obtained results from Table 1 with corresponding results ($PSQNR_{ex(32)}^{inf}$ and $PSQNR_{eq(64)}^{inf}$) from [14], the obtained experimental gain is shown in Table 2 for the same number of quantization levels.

Table 2 Experimental gain of the prposed model in
comparission to the piecewise uniform quantization model.

N	$\hat{\sigma}_d$	*GAIN*[dB]
	15	4.93
32	17	4.30
	29	1.87
	15	3.43
64	24	2.71
	29	0.36

By observing Table 2, it can be concluded that fixed non-uniform quantizers designed for discrete input samples for $N = 32$ and $N = 64$ quantization levels gives from 0.35605 to 4.93 [dB] higher *PSQNR* compared to the fixed piecewise uniform quantizers designed for discrete input samples

In addition, comparing theoretical results from Table 1 with $PSQNR_{th(32)}^{inf}$, it can be concluded that beside experimental gain, the proposed improved theoretical model that uses discrete variance predicts gain up to 4.52 [dB] compared to the same similar system, confirming experimental results.

5. CONCLUSION

In this paper we described a novel method for non-uniform quantizer design for discrete input samples and we tested the proposed quantizer for grayscale image coding. Considering that quantizers designed for continuous and discrete signals have different nature, we have introduced discrete designing variance as an additional and effective parameter in the

process of quantizer designing, for discrete input samples. System performance was discussed using weighting averaging of *PSQNR* for a set of three standard test grayscale images. The experimental results demonstrate that the performance of the proposed method outperforms other similar models - obtained gain of the proposed discrete solution is much higher for the most of discussed cases than the maximal difference of *PSQNR* between piecewise uniform (*L*=16) and optimal non-uniform quantizer that is equal to 0.7 [dB] (for continual input signal), which proves the introduction of the proposed quantizer design. Furthermore, additional system modification was proposed to adjust quantizer design in the special cases. However, this modification requires additional computing time as well as information about a set of input images. To generalize this approach, future work will include testing of specific images in order to find optimal values of input range support as well as implementation for different types of discrete input source.

REFERENCES

[1]   Jayant N. S., Noll P, *Digital Coding of Waveforms*, Prentice Hall Pb, 1984.

[2]   Yun Q., Shi, Huifnag Sun, *Image and Video Compression for Multimedia Engineering*, Taylor & Francis Group, 2008.

[3]   M. Savic, Z. Peric, N. Simic, "Coding algorithm for grayscale images based on Linear Prediction and dual mode quantization", *Expert Systems with Applications*, vol. 42, pp. 7285–7291, 2015.

[4]   N. Eslahi, A. Aghagolzadeh, "Compressive Sensing Image Restoration Using Adaptive Curvelet Thresholding and Nonlocal Sparse Regularization", *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3126 – 3140 July 2016.

[5]   J. Musić, T. Marasović, V. Papić, I. Orović, S. Stanković, "Performance of Compressive Sensing Image Reconstruction for Search and Rescue", *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, pp. 1739 – 1743, Nov. 2016.

[6]   A. Napieralski, J. Cłapa, K. Grabowski, M. Napieralska, W. Sankowski, P. Sękalski, M. Zubert, "Image and Video Processing with FPGA Support Used for Biometric as well as Other Applications", *Facta Universitatis, Series: Electronics and Energetics*, vol. 28, no. 2, June 2015, pp. 165 – 175.

[7]   Y. Yang, Q. Chen, Y. Wan, "A fast near-optimum block truncation coding method using a truncated K-means algorithm and intre-block correlation", *International Journal of Electronics and Communications (AEU)*, 2011, no. 65, pp. 576-581.

[8]   S. Kim, D. Lee, J-S. Kim, H-J. Lee, "A Block Truncation Coding Algorithm and Hardware Implementation Targeting 1/12 Compression for LCD Overdrive", *Journal of Display Technology*, vol. 12, no. 4, pp. 376−389, April 2016.

[9]   J-M., Guo, Y-F., Liu, "Improved Block Truncation Coding Using Optimized Dot Diffusion", *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp.1269−1275, March 2014.

[10] J-M., Guo, H. Prasetyo, N-J., Wang, "Effective Image Retrieval System Using Dot-Diffused Block Truncation Coding Features", *IEEE Transactions on Multimedia*, vol. 17, no. 9, pp. 1576−1590, September 2015.

[11] J-M., Guo, Y-F., Liu, "High Capacity Data Hiding for Error-Diffused Block Truncation Coding", *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4808−4818, December 2012.

[12] M. Savić, Z. Perić, M. Dinčić, "Design of Forward Adaptive Uniform Quantizer for Discrete Input Samples for Laplacian Source", *Electronics and Electrical Engineering*, no. 9 (105), pp. 73-76, 2010.

[13] M. Savić, Z. Perić, M. Dinčić, "An Algorithm for Grayscale Image Compression Based on the Forward Adaptive Quantizer Designed for Signals with Discrete Amplitudes", *Electronics and Electrical Engineering*, no. 2 (118), pp. 13-16, 2012.

[14] M. Savic, Z. Peric, M. Dincic, "Coding Algorithm for Grayscale Images Based on Piecewise Uniform Quantizers*", Informatica*, vol. 23, no. 1, pp. 125-140, 2012.

[15] Z. Peric, M. Petkovic, M. Dincic, "Simple Compression Algorithm for Memoryless Laplacian Source Based on the Optimal Companding Technique", *Informatica*, vol. 20, no. 1, pp. 99–114, 2009.

[16] Z. Peric, J. Nikolic, "An effective method for initialization of Lloyd-Max's algorithm of optimal scalar quantization for laplacian source", *Informatica*, vol. 18, no.2, pp. 279-288, 2007.

[17] N. Simic, Z. Peric, M. Savic, "Improved Algorithm for Grayscale Image Compression Based on Multimode Coding Algorithm", *Revue Roumaine des Sciences Techniques-Serie Electrotechnique et Energetique*, Tome 59, Issue 3, pp. 315-323, October 2014.

[18] Z. Peric, N. Simic, M. Savic, "Analysis and Design of Two Stage Mismatch Quantizer for Laplacian Source", *Elektronika ir Elektrotechnika*, vol. 21, no. 3, pp. 49-53, 2015.