Algorithm for optimizing quantization scales by an arbitrary quality measure

Nikolay Glumov
Department of GIS and ITsec
Samara National Research University;
MMIP laboratory
Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS
Samara, Russia
nglu@smr.ru

Mikhail Gashnikov
Department of GIS and ITsec
Samara National Research University;
MMIP laboratory
Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS
Samara, Russia
mih-fastt@yandex.ru

***Abstract*—This work deals with the task of constructing quantization scales optimal by an arbitrary criterion. Besides, these scales also satisfy the selected constraint. We consider a formal description of this optimization problem. We propose an algorithm for constructing quasi-optimal quantization scales that approximate optimal scales with the required precision, subject to the constraint. We formulate requirements for the optimization criterion and the restriction, ensuring the performance of the algorithm. We investigate the proposed quasi-optimal scales using computational experiments. The experimental results confirm the advantage of the constructed scales over the known ones.**

Keywords*—*quantization scale, non-uniform scale, quantization error, quantizer optimization, standard error

I. Introduction

Quantization [1] is the process of mapping input values from a large set to output values in a smaller set. In other words, quantization is rounding to a predetermined set of values.

You can use quantization to solve various problems: processing the phase space of the heart rhythm [2], normalizing the parameters of neural networks [3], embedding digital watermarks [4-5], processing spaces of semi-differentiable functions [6], and quantization of interpolation errors during compression [7-9], etc.

In this paper, we generalize the quantizer [10], proposed as part of the solution to the compression problem, for the case of an arbitrary quality measure and arbitrary restriction, which we use when optimizing the quantization scale. Besides, we also formulate requirements for this quality measure and this restriction. The fulfillment of these requirements allows us to ensure the performance of the optimization algorithm of the quantization scale.

II. The most common quantization scales

We describe the most common quantization scales. We divide the set of input values into quantization intervals. We specify the quantization level within each quantization interval. We call the quantization scale the set of the quantization levels and the quantization intervals.

Quantization means that we replace all input values belonging to the quantization interval with the corresponding quantization levels. Therefore, the quantization scale ultimately determines the quantization result. Most often, we use a uniform scale in which the intervals are the same size, and the levels are at the centers of the intervals.

However, the use of uniform scales in many situations leads to an unacceptably significant error, in particular with a small number of levels. In these cases, we use non-uniform scales. Lloyd-Max scales [7] are the most famous of the non-uniform quantization scales. We build these scales based on minimizing the root mean square error. The constraint is a given (fixed) number of quantization levels.

Despite the optimality of the error, the Lloyd-Max scales are not the best when solving many applied problems, since we often need to optimize not some error, but some other quality measure. Besides, the levels number of the quantization scale may not be known, which entails the need to use a different constraint when optimizing the scale, other than the constraint on the number of levels.

For example, we need to minimize the compressed data size with a fixed error in the compression problem [7-10]. We have to replace both the quality measure and the constraint to optimize the quantization scale for compression.

In this paper, we propose an algorithm for constructing a quantizer that is optimal by the required criterion. Besides, we take into account the required constraint when constructing this quantizer. We also describe the requirements for this criterion and the requirements for this constraint necessary for the operability of the proposed algorithm.

III. Quantizer optimization

A. Non-uniform scale quantizer

We describe a quantizer with a non-uniform [1, 8] scale. Let the input value $x \in (L, R]$ be an integer (for simplicity). We consider the range $(L, R]$ as the union of quantization intervals $(b_j, b_{j+1}]$:

$$(L, R] = \bigcup_{j=0}^{N-2} (b_j, b_{j+1}],\ b_j \in Z, b_0 = L-1,\ b_{N-1} = R, \tag{1}$$

where N is the number of boundaries b_j of the quantization intervals $(b_j, b_{j+1}]$. We denote $\mathbf{b}$ the vector of the boundaries of the quantization intervals:

$$\mathbf{b} = \left\{ b_j \in Z : b_j < b_{j+1}, b_0 = L-1, b_{N-1} = R,\ j \in [0, N) \right\} \tag{2}$$

We specify the quantization levels $c_j \in (b_j, b_{j+1}]$ within the intervals $(b_j, b_{j+1}]$. We denote $\mathbf{c}$ the vector of quantization levels:

$$\mathbf{c} = \left\{ c_j \in Z : b_j \le c_j \le b_{j+1},\ j \in [0, N-2] \right\} \tag{3}$$

The quantization level c_j belonging to the interval $(b_j, b_{j+1}]$ is the result of quantization of the input value x if x belongs to the interval $(b_j, b_{j+1}]$:

$$\Phi_q(x) = c_j : x \in (b_j, b_{j+1}], \tag{4}$$

where $\Phi_q(x)$ is the quantization function.

Requirement 1. Let $C(b_j, b_{j+1})$ there be a function that calculates the quantization level c_j for the corresponding interval $(b_j, b_{j+1}]$

$$c_j = C(b_j, b_{j+1}). \tag{5}$$

Then, to set the quantization scale, it suffices to specify the vector $\mathbf{b}$ of the boundaries of the quantization intervals (and the number *N* of components of this vector). Therefore, we use the terms "quantization scale $\mathbf{b}$" and "quantization scale $(b_0, ..., b_{N-1})$".

B. *Statement of the problem of quantizer optimization*

We denote $Q(\mathbf{b})$ the quality measure that we optimize when constructing a quantization scale. We perform this optimization using a constraint metric $E(\mathbf{b})$ that should not exceed the limit value $E_{\max}$.

We need to calculate the number of quantization intervals *N* and the boundaries $\mathbf{b}$ of the intervals to build the scale. Thus, we write the task of the scale optimization in the form:

$$\begin{cases} Q(\mathbf{b}) \to \min\limits_{N, \mathbf{b}} \\ E(\mathbf{b}) \le E_{\max} \end{cases} \tag{6}$$

Requirement 2. There must be a way to calculate the quality measure of the scale $(b_0, ..., b_{N-1})$ through the quality measures of the subscale $(b_0, ..., b_{N-2})$ and the subscale (b_{N-2}, b_{N-1}). For simplicity, we further assume that we can simply summarize the quality measures of such scales (we can use this algorithm also for more sophisticated ways of calculating quality measures):

$$Q(b_0, ..., b_{N-1}) = Q(b_0, ..., b_{N-2}) + Q(b_{N-2}, b_{N-1}). \tag{7}$$

Requirement 3. Let the similar requirement also be true for the constraint (the ability to calculate the constraint for the scale through the corresponding subscales):

$$E(b_0, ..., b_{N-1}) = E(b_0, ..., b_{N-2}) + E(b_{N-2}, b_{N-1}). \tag{8}$$

We can see that the formulated requirements are quite weak, as they are true in most practical situations.

C. *The forward procedure of the quantizer optimization algorithm*

Let Δ_E be a small step in the value of the constraint metric (algorithm parameter). We split the range $[0..E_{\max}]$ of the constraint metric into K sub-ranges of the equal sizes $\Delta_E = E_{\max} / K$.

Step number 1. Construction of optimal scales of two quantization intervals.

We build optimal scales at all intervals $[L, r],\ r \in [L+1, R]$. These scales must satisfy the following conditions:

a) The scale consists of two quantization intervals.

b) Scale constraint metric $E(\mathbf{b}) = k\Delta_E,\ k \in [0, K-1]$.

We write the quantization intervals of these scales in the form:

$$(b_0, b_1] = (L, d],\ (b_1, b_2] = (d, R],\ L \le d < r, \tag{9}$$

where *d* is the only interval boundary that we need to choose for each desired scale.

We can write the quality measure of the scale of two intervals through the quality measure at intervals:

$$q^{(1)}(r, d) = Q(L, d) + Q(d, r). \tag{10}$$

We can write the constraint metric of the scale of two intervals similarly:

$$e^{(1)}(r, d) = E(L, d) + E(d, r). \tag{11}$$

Then we can put the optimal value of the boundary $d \in [L, r)$ in the matrix $B^{(1)}$:

$$\begin{gathered} B^{(1)}_{r,k} = \arg\min_{L \le d < r} \left\{ q^{(1)}(r, d) : e^{(1)}(r, d) \le k\Delta_E \right\}, \\ r \in [L+1, R], k \in [0, K-1] \end{gathered} \tag{12}$$

Each element $B^{(1)}_{r,k}$ contains the boundary $d \in [L, r)$ of the scale of two intervals. This scale is optimal in the range $[L, r],\ r \in [L+1, R]$. The constraint metric of this scale $E(\mathbf{b}) = k\Delta_E,\ k \in [0, K-1]$.

Besides, we put the corresponding values of the scale quality measure in the matrix $Q^{(1)}$:

$$Q^{(1)}_{r,k} = q^{(1)}(r, B^{(1)}_{r,k}), r \in [L+1, R], k \in [0, K-1].$$

Step number *j*. Construction of optimal scales of $(j+1)$ intervals.

We build optimal scales at all intervals $[L, r],\ r \in [L+1, R]$. These scales must satisfy the following conditions:

a) The scale consists of $(j+1)$ quantization intervals.

b) Scale constraint metric $E(\mathbf{b}) = k\Delta_E, k \in [0, K-1]$.

We search through the penultimate boundary $d \in [L + j - 1, r)$ of each of these scales.

We found all the optimal scales of j quantization intervals during the previous step of the algorithm. These scales are optimal at intervals $[L, d]$, $d \in [L + j - 1, r)$. The constraint metric of these scales is equal $E(\mathbf{b}) = k\Delta_E - E(d, r)$. Then we can consider scales that satisfy the following conditions:

a) The scale is in the interval $[L, r]$, $r \in [L+1, R]$.

b) The scale consists of $(j+1)$ quantization intervals.

c) The scale contains the optimal subscale of j intervals.

d) Scale constraint metric $E(\mathbf{b}) = k\Delta_E, k \in [0, K-1]$.

We can write the quality measure of all these scales in the following form:

$$q^{(j)}(r, k, d) = Q^{(j-1)}_{d,t} + q(d, r),\ t = \left\lceil \frac{(k\Delta_E - E(d, r))}{\Delta_E} \right\rceil \tag{13}$$

Optimization of this function by d allows us to calculate the penultimate boundary of the desired optimal scale. We put this boundary in the matrix $B^{(j)}$:

$$B^{(j)}_{r,k} = \arg\min_{L \le d < r} q^{(j)}(r, k, d),\ r \in [L+1, R], k \in [0, K-1] \tag{14}$$

We also put the appropriate quality measure in the matrix $Q^{(j)}$:

$$Q^{(j)}_{r,k} = q^{(j)}(r, B^{(j)}_{r,k}), r \in [L+1, R], k \in [0, K-1] \tag{15}$$

The forward procedure of the quantizer optimization algorithm stops at step number $R - L - 1$. Then the reverse procedure of the quantizer optimization algorithm starts.

D. *The reverse procedure of the quantizer optimization algorithm*

The one-dimensional array $\tilde{Q}(j) = Q^{(j)}_{R,K-1}, 0 < j < R - L$ contains the quality measure of the scales of $(j+1)$ quantization intervals. These scales are optimal in the range $(L, R]$ with the constraint metric $E(\mathbf{b}) = (K-1)\Delta_E$.

The minimum by j in the array $\tilde{Q}(j)$ corresponds to the step number at which we built the desired optimal scale. The number of quantization levels of this scale is two more than this step number:

$$N = 2 + \arg\min_{0<j<R-L} \tilde{Q}(j) = 2 + \arg\min_{0<j<R-L} Q^{(j)}_{R,K-1} \tag{16}$$

We know the first $b_0 = L$ and last $b_{N-1} = R$ boundaries of the quantization intervals of this optimal scale. We get the penultimate boundary b_{N-2} of this scale from the matrix B:

$$b_{N-2} = B^{(N-2)}_{R,K-1} \tag{17}$$

We calculate the boundaries of the remaining intervals of the optimal scale using the following recursive procedure:

$$b_j = B^{(j)}_{v,w}, v = b_{j+1}, w = \left\lceil \left(E_{\max} - \sum_{i=j+1}^{N-2} E(b_i, b_{i+1}) \right) \Big/ \Delta_E \right\rceil \tag{18}$$
$$j = N-3, N-2, \ldots, 1$$

This completes the construction of the scale. The constructed scale is quasi-optimal (asymptotically optimal for Δ_E tending to zero). The constraint metric of the constructed scale is in the range $[E_{\max} - N\Delta_E, E_{\max}]$ since the deviation of the constraint metric of the constructed scales from the constraint metric of the optimal scale increases no more than Δ_E at each step of the algorithm.

IV. Experimental Study of the Quantizer Optimization Algorithm

We performed computational experiments to study the effectiveness of the proposed quantizer optimization algorithm. We investigated the problem of constructing quantization scales that are optimal for the compression problem with a controlled error [7-12]. These scales allow us to minimize the compressed size data with a fixed error.

We used the quality measure equal to the entropy $H(\mathbf{b})$ [7] of the quantized values since this entropy approximates well the compressed data size. We also used the scale constraint metric equal to the variance $\varepsilon^2_{MSE}(\mathbf{b})$ of the quantization error [12]. Therefore, we solved the optimization problem (6) in the form:

$$\begin{cases} Q(\mathbf{b}) = H(\mathbf{b}) = -\sum_{j=0}^{N-2} P(b_j, b_{j+1}) \log_2 P(b_j, b_{j+1}) \to \min_{N,\mathbf{b}} \\ E(\mathbf{b}) = \varepsilon^2_{MSE}(\mathbf{b}) = \sum_{j=0}^{N-2} \sum_{x=b_j}^{b_{j+1}} f(x)\left(x - C(b_j, b_{j+1})\right)^2 \le \varepsilon^2_{\max} \end{cases} \tag{19}$$

Here $f(x)$ is the probability density of the input value x equal to the interpolation error of the compressible signal samples.

We describe the probability $P(l, r)$ of falling x into the interval $(l, r]$ and the function $C(l, r)$ of calculating the quantization level from the quantization interval (l, r) as follows:

$$P(l, r) = \sum_{x=l+1}^{r} f(x),\ C(l, r) = \sum_{x=l+1}^{r} x f(x) \Big/ \sum_{x=l+1}^{r} f(x) \tag{20}$$

Here, the quantization level $C(l, r)$ is equal to the local average over the quantization interval (l, r).

We used the distribution density $f(x) = \exp(-\xi |x|)/(2\xi)$. This type of distribution density is natural for the interpolation error of differential [11], hierarchical [9], and many other [7, 12] compression methods.

We used uniform quantization scales and non-uniform Lloyd-Max scales [7] as a basis for comparison. We built the Lloyd-Max scales based on minimizing the root mean square (RMS) error $\varepsilon_{\Sigma}^{2}(\mathbf{b})$ while limiting the number of quantization levels $N = N_{\max}$:

$$\begin{cases} \varepsilon_{\Sigma}^{2}(\mathbf{b}) = \sum_{j=0}^{N-2} \sum_{x=b_j}^{b_{j+1}} f(x)\left(x - C\left(b_j, b_{j+1}\right)\right)^2 \to \min_{\mathbf{b}} \\ N = N_{\max} \end{cases} \tag{21}$$

We show a graph of the dependence of the entropy of quantized interpolation errors on the RMS interpolation error in Fig. 1. You can see that the proposed algorithm allows you to build scales that have an advantage in the “error-entropy” coordinates over uniform scales and Lloyd-Max scales.

The proposed quantizer provides a smaller compressed data size with the same error. Accordingly, this quantizer provides a smaller error with the same compressed data size. This advantage increases with the increase in the quantization error.

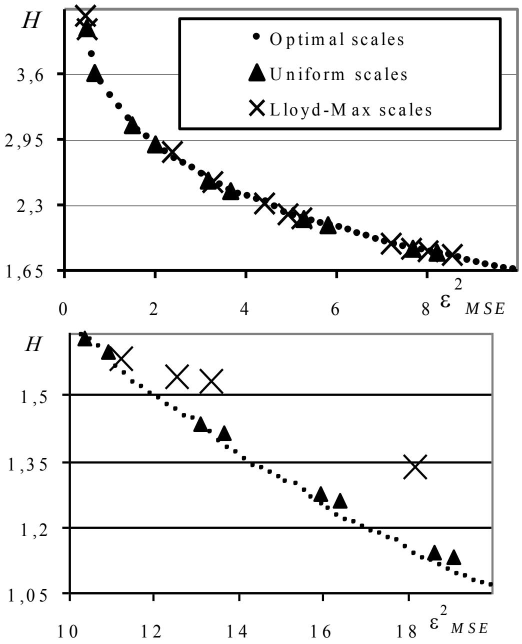

Fig. 1. The study of the algorithm for constructing optimal quantization scales.

V. CONCLUSION

We considered the problem of constructing quantization scales that are optimal by a given criterion and satisfy the chosen constraint. We considered the precise formulation of such an optimization problem. We proposed an algorithm for constructing quasi-optimal quantization scales approximating optimal scales with a given precision, subject to the constraint.

We formulated requirements for the optimization criterion and the constraint, ensuring the operability of the optimization algorithm of the quantization scale. We performed computational experiments to construct quasi-optimal scales. The obtained experimental results confirmed the advantage of the constructed quantization scales over the known ones.

ACKNOWLEDGMENT

The reported study was funded by RFBR according to the research projects 18-01-00667 (sections II, III.B, III.C, III.D, IV, V), 18-07-01312 (section III.A), and the RF Ministry of Science and Higher Education within the state project of FSRC “Crystallography and Photonics” RAS under agreement 007-GZ/Ch3363/26 (section I).

REFERENCES

[1] A. Oppenheim and R. Schafer, “Digital Signal Processing,” Litres, 2018, 1048 p.

[2] A. Kudinov, S. Mikheev, V. Tsvetkov and I. Tsvetkov, “Quantization of the phase space of an instantaneous heart rate,” International conference “Mathematical biology and bioinformatics”, vol. 7, e15, 2019.

[3] D. Kolesnikov and D. Kuznetsov, “Optimal uniform quantization of parameters of convolutional neural networks,” Questions of radio electronics, vol. 8, pp . 99-103, 2018.

[4] O. Evsyutin, A. Kokurina and R. Meshcheryakov, “Steganographic embedding of additional data in Earth remote sensing images using the QIM method with a variable quantization step in the frequency domain,” Bulletin of the Tomsk polytechnic university of engineering georesources, vol. 330, no. 8, pp. 155-162, 2019.

[5] A. Egorova and V. Fedoseev, “A classification of semi-fragile watermarking systems for JPEG images,” Computer Optics, vol. 43, no. 3, pp. 419-433, 2019. DOI: 10.18287/2412-6179-2019-43-3-419-433.

[6] A. G. Sergeev, “Quantization of the Sobolev space of half-differentiable functions II,” Russian Journal of Mathematical Physics, vol. 26, no. 3, pp. 401-405, 2019.

[7] W. Pearlman and A. Said, “Digital signal compression: principles and practice,” Cambridge university press, 2011, 419 p.

[8] V. Soifer, “Computer image processing, Part II: Methods and algorithms,” Saarbrücken: VDM Verlag, 2010, 584 p.

[9] V. Sergeyev, N. Glumov and M. Gashnikov, “Compression method for real-time systems of remote sensing,” 15th international conference on pattern recognition, vol. 3, pp. 232-235, 2000.

[10] M. Gashnikov, “Optimal quantization in the task of digital signal compression,” Computer Optics, vol. 21, pp. 179-184, 2001.

[11] A. Maksimov and M. Gashnikov, “Adaptive interpolation of multidimensional signals for differential compression,” Computer Optics, vol. 42, no. 4, pp. 679-687, 2018. DOI: 10.18287/2412-6179-2018-42-4-679-68.

[12] K. Sayood, “Introduction to data compression,” The Morgan Kaufmann series in multimedia information and systems, 2012, 743 p.