



Implementation of perceptual measure Picture Quality Scale with neural network to evaluate distortions in compressed images

Maria Skublewska-Paszkowska*, Jakub Smolka†

*Computer Science Institute, Lublin University of Technology,
ul. Nadbystrzycka 36b, 20-618 Lublin, Poland*

Abstract

The perceptual measures are often used to assess distortions in image compression. In this article different images were evaluated using the Picture Quality Scale (PQS) measure with neural network. On the basis of original and compressed images the local distortions in the compressed image are calculated. Then five factors $\{F_1, F_2, F_3, F_4, F_5\}$, which represent these distortions are computed, and used to evaluate correlations among them by the covariance matrix. The new values are put to the input of neural network, to calculate the single PQS value. During the process of learning the neural network the best value PQS, which reflects the largest degree of particular distortions in the compressed images is obtained. The images are divided into three groups: faces, landscapes and shapes. The process of learning is controlled by the subjective measure Mean Opinion Score (MOS) with 15 observers.

1. Introduction

Nowadays not only differences between the original and compressed images such as MSE or PSNR are computed but also the visual differences between

*e-mail address: mariap@pluton.pol.lublin.pl

†e-mail address: jsmolka@pollub.pl

those images are taken into the account. One of the perceptual measures is Picture Quality Scale (PQS). It is used to assess the visual distortions in compressed images on the basis of Mean Opinion Score (MOS) measure. The usage of PQS measure together with neural network enables easy assessment of different distortions without observers. In this article the assessment of compressed images with the Daubechies wavelet of degrees: 4, 6, 8, 14 and 20 is presented. The article is divided into three parts. Firstly, the PQS measure is described. Secondly, the scheme of neural network is presented. Finally, the research as well as the process of learning of neural network and the obtained results are shown.

2. Picture Quality Scale Measure

The Picture Quality Scale (PQS) [1, 2] is an objective measure that is used to evaluate perceptual distortions in the compressed image using the original one. The PQS schema is presented in Fig. 1. It involves the Human Visual System features and subjective measure Mean Opinion Score (MOS) [1, 2]. In the PQS measure five different local distortion maps f_1, f_2, f_3, f_4, f_5 in the compressed image are calculated. They are used to the gain distortion vector $F = \{F_1, F_2, F_3, F_4, F_5\}$. This vector is necessary to compute the covariance matrix and then both the eigenvalues and eigenvector of values are calculated. The distortion parameters $Z = \{Z_1, Z_2, Z_3, Z_4, Z_5\}$ are acquired from the eigenvector and distortion vector F .

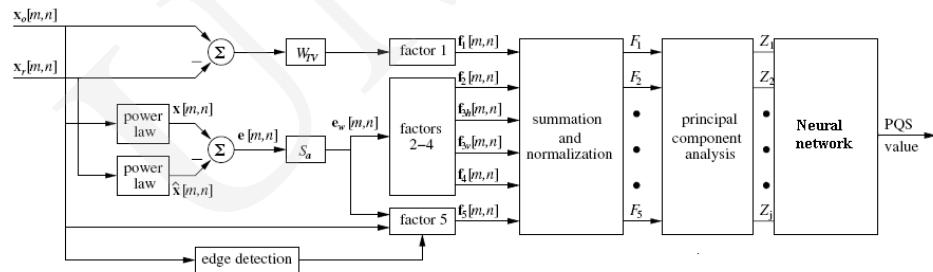


Fig. 1. The PQS schema [1] with neural network

To point out the PQS value two images are necessary: the original one x_0 and the compressed one x_r . The matrix f_1 (1) [1, 2] represents the difference between the original and reconstructed images convolved with frequency weighting W_{TV} [1].

$$f_1[m, n] = ((x_0[m, n] - x_r[m, n]) * W_{TV}[m, n])^2. \quad (1)$$

The matrix f_2 (2) [1, 2] reflects the distortions between the original and the compressed images which are adapted by the Werner Fechner law [1] and the spatial domain counterpart of the frequency response $S_a(\omega)$ [1]. Additionally, this distortion cuts the values which are below thresholds T by an indicator function I_T [1, 2].

$$f_2[m, n] = I_T(e[m, n] * S_a[m, n])^2 \quad (2)$$

$$e[m, n] = (x_0[m, n] - x_r[m, n]). \quad (3)$$

The matrix f_3 (4) [1], [2] illustrates the differences in linear elements of images, especially the end of block disturbances [1, 2]. This matrix is made of two components that is the error in vertical and horizontal block discontinuities. The equations for the vertical matrix are analogical to those for the horizontal matrix (4) and (5) [1, 2]. The matrix e_w represents the convolution between the matrix e (3) and the modelled contrast sensitivity function (S_a) [1, 2].

$$f_{3h}[m, n] = I_H[m, n] \Delta_h^2[m, n] \quad (4)$$

$$\Delta_h[m, n] = e_w[m, n] - e_w[m, n + 1] \quad (5)$$

In the compressed images the correlation distortions can be often seen. They can be present in texture, image features and blocking effects [1, 2]. The local distortion map f_4 (6) [1, 2] is evaluated with the use of local spatial correlations.

$$f_4[m, n] = \sum_{(k,l) \in W; l=k \neq 0} |r(m, n, k, l)|^{0.25} \quad (6)$$

$$r(m, n, k, l) = \frac{1}{W-1} \left[\sum_{i,j} e_w[i, j] e_w[i+k, j+l] - \frac{1}{W} \sum_{i,j} e_w[m, n] \sum_{i,j} e_w[i+k, j+l] \right]$$

The parameter W represents the number of elements in the window of five by five items. Additionally $|l| \leq 2$ and $|k| \leq 2$ must be fulfilled [1].

The last distortion that is taken into account in the PQS measure is the vicinity of high contrast image transitions [1, 2]. The two Human Visual System features are used in that item that are visual masking and enhanced visibility of misalignments. The f_5 distortion is defined as [1, 2]:

$$f_5[m, n] = I_M[m, n] |e_w[m, n]| (S_h[m, n] + S_v[m, n]).$$

The I_M is an indicator function that selects only the items close to high intensity transitions. The S_h and S_v stand for the horizontal and vertical masking factors

respectively (7) [1, 2].

$$S_h[m, n] = e^{\{-0.04V_h[m, n]\}} \quad (7)$$

$$V_h[m, n] = \frac{|x_0[m, n - 1] - x_0[m, n + 1]|}{2}.$$

The distortion factors are based on the local distortion maps (from f_1 to f_5). The first factor is the sum of all items from the matrix f_1 divided by the sum of squared items from the original image (8) [1, 2].

$$F_1 = \frac{\sum_m \sum_n f_1[m, n]}{\sum_m \sum_n x_0^2[m, n]}. \quad (8)$$

The second distortion is computed as the sum of all matrix f_2 elements divided by the sum of squared values of compressed image (9) [1, 2].

$$F_2 = \frac{\sum_m \sum_n f_2[m, n]}{\sum_m \sum_n x_r^2[m, n]}. \quad (9)$$

The third factor is the square of two elements: the block error discontinuities at vertical F_{3v} and horizontal F_{3h} block boundaries (10) [1, 2].

$$F_3 = \sqrt{F_{3h}^2 + F_{3v}^2} \quad (10)$$

$$F_{3h} = \frac{1}{N_h} \sum_m \sum_n f_{3h}[m, n]$$

$$F_{3v} = \frac{1}{N_v} \sum_m \sum_n f_{3v}[m, n].$$

The parameters N_h and N_v stand for the number of items pointed out by an indicator function. The fourth factor is the sum of items of matrix f_4 divided by the size of images, where N represents the number of columns and M – the number of rows (11) [1, 2].

$$F_4 = \frac{1}{MN} \sum_m \sum_n f_4[m, n]. \quad (11)$$

The last factor is the sum of all elements of matrix f_5 divided by the parameter that represents the number of points, whose values after the Kirsch operator, are greater or equal to 400 (12) [1, 2].

$$F_5 = \frac{1}{N_k} \sum_m \sum_n f_5[m, n]. \quad (12)$$

On the basis of the distortion factors vector the covariance matrix is computed and normalized (13) [1, 2, 3] so that the values are from the [0, 1] interval.

$$\text{Cov}[f_i, f_j] = \frac{\sum_m \sum_n |f_i[m, n] - \bar{f}_i| \sum_m \sum_n |f_i[m, n] - \bar{f}_j|}{\sqrt{\sum_m \sum_n (f_i[m, n] - \bar{f}_i)^2 \sum_m \sum_n (f_i[m, n] - \bar{f}_j)^2}}. \quad (13)$$

The parameter \bar{f}_i represents the average value of matrix f_i .

On the basis of the covariance matrix the eigenvalues and eigenvector are computed [1]. The vector Z is calculated as the multiplication of covariance matrix by the factor values.

The PQS value should be similar to the Mean Opinion Score (MOS) measure [1, 2]. This measure is based on the assessments of visual distortions in images by the group of observers (14) where M represents the number of observers and $A(i, k)$ stands for the assessment of i^{th} observers of k^{th} image. The observers are obliged to adjust the value to the distortions that they notice. Possible assessment values are shown in Table 1 [1, 2].

Table 1. Possible assessment values for the MOS measure

Assessment value	Visual distortions
1	The image is illegible
2	The distortions are very visible
3	The distortions are visible
4	The distortions are slightly visible
5	The distortions are invisible

$$MOS = \frac{1}{M} \sum_{i=1}^M A(i, k). \quad (14)$$

3. Neural network

In order to make the implementation of Picture Quality Scale without the observers easier, the feed-forward neural network [4] was adapted. The network consists of three layers: input layer, hidden layer and output layer. The first layer is made of five neurons all of which have five inputs and one output. The second layer is made of 9 neurons all of which possess five inputs and one output. The last one has got only one neuron with nine inputs and one output that generates the output of the whole network. This output is the PQS value. The scheme of the used network is presented in Fig. 2. The training function implemented in this network to update the weight and bias the values is the

Levenberg–Marquardt algorithm [5]. The neurons in the first and second layers have the hyperbolic tangent transfer function and the final neuron possesses the linear one so that the output of neural network does not have to be scaled.

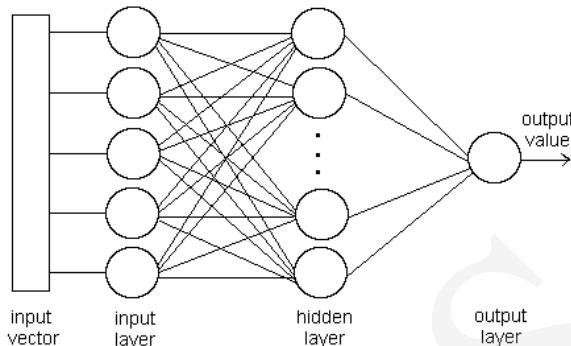


Fig. 2. The scheme of feed-forward network

4. The research

The Z vectors were computed for the squared colour images of the size 512 by 512 pixels.

The evaluated images were compressed with the wavelet transform Daubechies for different degrees: four, six, eight, fourteen and twenty [6, 7]. EZW (*Embedded Zerotree Wavelet*) [8, 9] was used to compress images. The compression rates were from 2 to 20.

There was a conversion from the RGB space color to luminance and chrominance spaces. For each space the vector Z is calculated. For the output Z vector the average is taken. These Z vectors are the input of neural network which calculates the PQS value. The images are assembled into three groups i.e. face, scenery and shape. There were sixty images with different compression rates from 2 to 20.

Each pair of original and compressed images was evaluated by fifteen observers so that the Mean Opinion Square measure (14) was calculated. For the evaluation of visual distortions in images the five-level scale was used (Table 1). The acquired data were doubled, but the sequence of input data was changed so that the neural network would learn not memorize the test data. These data were put as the target for the neural network for the process of learning. The training graph is presented in Fig. 3.

The error for evaluating the distortions in the compressed images according to the original one is 0.209338. The network results for the data from the learning set are good. The values from neural network are similar to the MOS

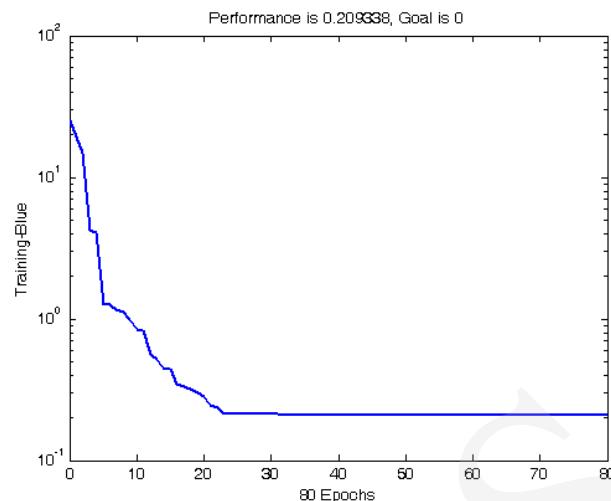


Fig. 3. The training graph for feed-forward neural network

values. The results are presented in Figures from 4 to 6 for the images divided into three groups: face, shape and landscape pictures respectively. The required values are mostly the same or slightly different from the MOS values. The PQS values for the new Z vectors are mostly correct according to the MOS measure, however, an error can occur. The results from the neural network and the MOS value for the new Z vectors, which were not included in the learning set, are presented in Fig. 7.

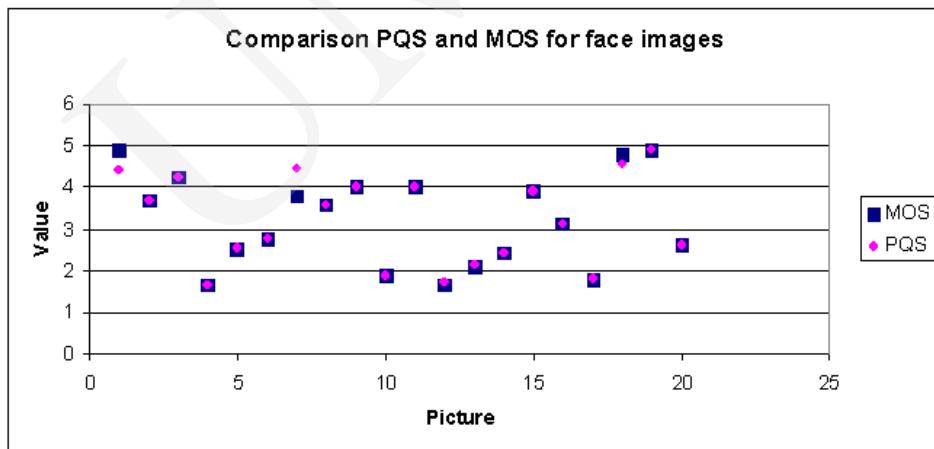


Fig. 4. The chart of the MOS values and the results obtained from neural network for face images

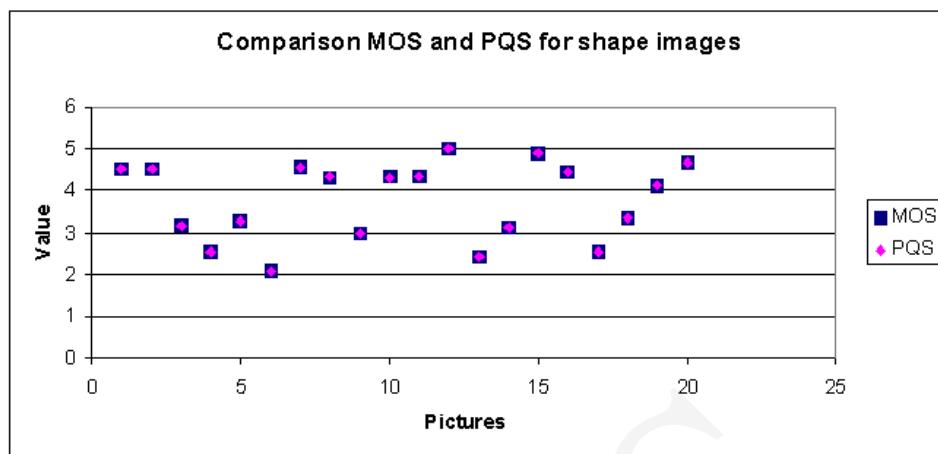


Fig. 5. The chart of the MOS values and the results obtained from neural network for shape images

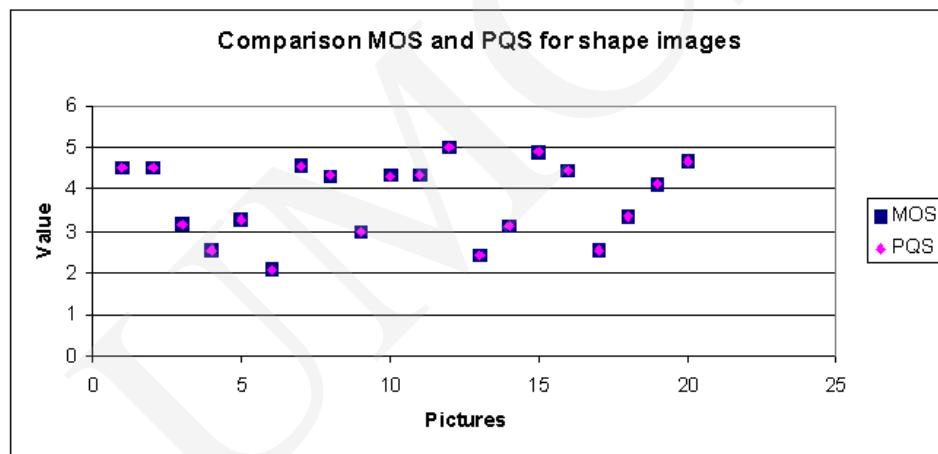


Fig. 6. The chart of the MOS values and the results obtained from neural network for landscape images

The exemplary evaluation of distortions from the learning set of the compressed image bardowl.bmp (Fig. 6) for 16.6 compression rate according to the original one was set to 1.888889 by the MOS measure while the neural network gave the 1.9937 value. After training the network, the new set of data was put into the input of neural network. The data were different from the learning data. The results obtained from the new set of data were also satisfactory. They are presented in Fig. 8. The maximum error was set to 1.321.



Fig. 7. The original (on the left) and compressed (on the right) images

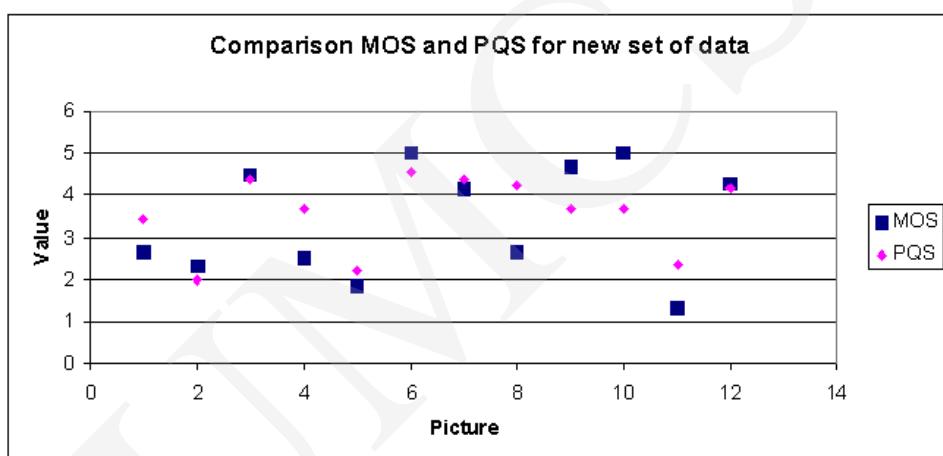


Fig. 8. Comparison of MOS and PQS for the new set of data

5. Conclusions

The use of neural network enables the computation of PQS value without observers so that the implementation of Picture Quality Scale is much easier. The neural network, which is presented in this article, generates mostly the correct values for different distortions of images compressed with the Daubechies wavelet transforms. The MOS measure is objective depending on the observers, their moods and conditions. That is why the maximum difference between the PQS and MOS values, which is set to 1.3211, is satisfactory. The idea of matching the visual distortions by the neural network to the MOS value is obtained.

The properly learned neural network is the perfect tool for evaluation of visual distortions in compressed images.

References

- [1] Wu H. R., Rao K. R., *Digital Video Image Quality and Perceptual Coding*, Taylor & Francis Group, London, New York, (2006).
- [2] Miyahara M., Kotani K. and Algazi V. R., *Objective Picture Quality Scale (PQS) for image coding*, IEEE Transactions on Communications, 46(9) (1998).
- [3] Cyganek B., *Komputerowe przetwarzanie obrazów trójwymiarowych*, EXIT (2002).
- [4] Tadeusiewicz R., Gąciarz T., Borowik B., Leper B., *Odkrywanie właściwości sieci neuronowych przy użyciu programów w języku C#* (2007).
- [5] Lourakis M. I. A., *A brief description of the Levenberg-Marquardt algorithm by levmar*, February 11 2005, <http://www.ics.forth.gr/lourakis/levmar/levmar.pdf>
- [6] Daubechies I., *Ten lectures on wavelets*, Rutgers University and AT&T Bell Laboratories.
- [7] Walker J. S., *A Primer on Wavelets and their Scientific Applications*, CHAPMAN & HALL/CRC.
- [8] Shapiro J. M., *Embedded Image Coding Using Zerotrees of Wavelet Coefficients*, IEEE Transactions on signal processing 41(I2) (1991).
- [9] EZW encoding, Valens 1999-2004,
<http://pagesperso-orange.fr/polyvalens/clemens/ezw/ezw.html#section2>