https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-260

Image Processing: Algorithms and Systems XVII
(IS&T, Springfield, VA, 2019)

Combined no-reference IQA metric and its performance analysis

Oleg Ieremeiev[a], Vladimir Lukin[a], Nikolay Ponomarenko[b], Karen Egiazarian[b]
[a] National Aerospace University, 61070, Kharkov, Ukraine;
[b] Tampere University of Technology, FIN 33101, Tampere, Finland

Abstract

The problem of increasing efficiency of blind image quality assessment is considered. No-reference image quality metrics both independently and as components of complex image processing systems are employed in various application areas where images are the main carriers of information. Meanwhile, existing no-reference metrics have a significant drawback characterized by a low adequacy to image perception by human visual system (HVS). Many well-known no-reference metrics are analyzed in our paper for several image databases. A method of combining several no-reference metrics based on artificial neural networks is proposed based on multi-database verification approach. The effectiveness of the proposed approach is confirmed by extensive experiments.

Keywords: image visual quality assessment, full-reference metrics, combined metrics, robust metrics

Introduction

Imaging systems are employed nowadays for numerous applications [1,2] where mass user segment (digital cameras, smart phones and so on) is the most known [2]. They produce a wide variety of images that are of different visual quality due to many factors: quality of an imaging system, conditions of imaging, principle of operation, etc. In this regard, a problem of automating analysis and enhancement of imaging data arises in digital image processing [3, 4, 5].

The first and decisive stage of such automatic systems is image quality assessment (IQA). As it is known, there are full reference, reduced reference and no-reference image visual quality assessment and various metrics used for this purpose [4]. In our case, when an image is acquired and there is no reference, one has to deal with a no-reference quality assessment. Due to this, there is a growing need for an adequate, accurate and reliable no-reference visual quality metric that, based on the characteristics of an image itself, evaluate its visual quality [5].

Currently, for objective reasons, the existing metrics do not provide appropriately high accuracy and stability [4]. The main reasons are the following. Visual quality assessment requires consideration of features of human visual systems (HVS), but this area of research is extremely complex and it is difficult to take into consideration all peculiarities of HVS. Therefore, the metrics are usually based on fragmented and simplified models of visual perception. In addition, a typical requirement to no-reference metrics is to provide a high computational efficiency. This imposes restrictions on the complexity of the mathematical models used and limits practical use of many already proposed metrics. Hence, efforts to design new or to modify existing no-reference visual quality metrics continue.

One approach to solve the problem of no-reference metric's design is to combine them in one or another way. This takes place for both full reference [6-8] and no-reference [9] metrics. There are different ways to combine metrics (that we will further call elementary). One way is to use some function of two or more elementary metrics and to optimize its parameters, for example, weights [10]. Another way, very popular nowadays, is to use pre-trained neural networks [11] including those ones based on deep learning [12]. The latter, trained on millions of images in the process of multi-parameter optimization of millions of neurons, form complex relationships. During their use, all analyzed images are divided into small fragments, which are classified according to their characteristics. After that, all parameters can be reduced to a single coefficient, which also characterizes the visual quality of the image. Deep learning has the advantages that it does not require mathematical models of HVS and metrics (similar or even more complex dependencies are formed during training). At the moment, this approach is too resource-intensive, especially for portable and low power devices. Besides, it requires intensive learning process based on collected opinions of humans.

Any approach based on learning (training) possesses one drawback. It might produce good results for data used in training or similar ones. However, it's effectiveness reduces immediately if there are differences in training and verification data. Thus, both training and verification should be done carefully.

Taking into account all these factors, this paper considers different methods of creating combined no-reference metrics and proposes a method of combining metrics based on neural networks that is acceptable for various automated systems of image processing. In addition, we pay attention to metric's verification using different sets of test images.

Image databases and metrics analysis

The mandatory step of metrics design and analysis is testing them on a specific set of images and comparing the obtained results with subjective quality assessments. These might be small sets of images for specific tasks. But a generally accepted approach is to use specialized image databases such as TID2008 [13], TID2013 [14], LIVE [15], and others [16-19]. Such test sets contain up to several thousand distorted images for which visual quality estimates (mean opinion scores, MOS) are determined based on the results of a large number of subjective experiments.

Note that image databases differ significantly in their characteristics. This determines both the tasks that can be realized on their basis and the effectiveness of the obtained solutions. A favorable distinction of the class of no-reference quality metrics is the possibility to use full-reference image databases. There are following key indicators that should be considered when using different image sets:

- Number of test images: full-reference image database TID2013 contains 3000 distorted images, the no-reference database KonIQ10k [16] contains more than 10,000 images.
- The number of various distortion types (the current maximum values are 17 for TID2008 and 24 for TID2013).
- Testing methodology and accuracy of the results of subjective experiments.
- The number of experiments and the final accuracy of the obtained MOS.

The analysis of different databases according to these characteristics (criteria) has been carried out in [13,14] and a more in-depth analysis has been done in [20] during development of TID2013.

The following image databases are considered in our paper: TID2013, NRTID [17], Live, Live MD [18], MDID [19], and KonIQ10k. The values of the Spearman rank order correlation coefficients (SROCC) of many open access no-reference metrics [21-31] for aforementioned databases are presented in Table 1.

Table 1. SROCC values of no-reference quality metrics

Metrics	Databases						Time sec
	TID 2013	NRTID	LIVE	LIVE MD	MDID	KonIQ 10k	
BIQI [21]	0.405	0.393	0.883	0.521	0.628	0.511	5,5
BLIINDS2 [22]	0.395	0.199	0.921	0.181	0.178	0.018	221
BRISQUE [23]	0.367	0.376	0.941	0.502	0.404	0.223	1,1
CDIIVINE [24]	0.373	0.597	0.958	0.235	0.489	0.491	72
DESIQUE [25]	0.069	0.157	0.404	0.199	0.056	0.078	1,5
DIIVINE [26]	0.344	0.492	0.817	0.660	0.532	0.431	53
FISH [27]	0.052	0.595	0.375	0.300	0.267	0.608	2,8
ILNIQE [28]	0.492	0.403	0.854	0.877	0.694	0.488	17
NIQE [29]	0.313	0.095	0.026	0.774	0.654	0.530	1,7
NJQA [30]	0.100	0.307	0.303	0.003	0.021	0.083	46,6
SMETRIC [31]	0.097	0.710	0.540	0.194	0.297	0.613	4,4

The above mentioned features and differences in the accuracy of given image databases are the reasons for significant differences in the correlation coefficients of each metric for the considered sets of images (databases). For example, the database LIVE has high MOS dispersion rate [13], which indicates a lack of MOS accuracy, and five most common and studied types of distortion. Therefore, this set is simple for quality metrics. As a result, high performance on LIVE (0.9 and above) does not guarantee the quality of the metric, since on other sets their accuracy may not exceed 0.2.

Analyzing the no-reference quality metrics themselves, we note that, except LIVE, their performance indicators practically do not exceed the range 0.6…0.7. Such values are quite low and insufficient for practical application of quality metrics. This is the reason for the lack and relevance of automated visual quality analysis systems based on IQA.

The second significant drawback of some of no-reference IQA is their computational complexity. In Table 1, the last column displays computational time for metrics for test images with the size of 1280 × 720 pixels from LIVE MD on a desktop computer with 2.5 GHz processor. One can see from these numbers that it is unfeasible to calculate these metrics for larger images on low power mobile devices.

Simple approaches to combine no-reference metrics

One way to improve metrics’ performance is to combine them [6-11]. This can be done using simple functions or operation with elementary metrics used as the arguments [10]. Examples of such methods for reference quality metrics are considered, in particular, in [6, 32].

In the paper [6], a resulting metric was obtained as a result of robust estimation applied to transformed elementary metrics. Optimization procedure presumes finding the best metrics (their sets) among available ones. Elementary metrics are fitted using MOS values and then linearized to avoid problems with different ranges of metric variation as well as peculiarities of behavior of elementary metrics.

In the second paper by Okarma [10], a combined metric of two metrics is defined in a simple way as

$$M_{Multiplied} = M_1^a \times M_2^b \tag{1}$$

where M1, M2 are some elementary metrics and the function parameters *a* and *b* are optimized to provide maximum of SROCC. Similarly, it is also possible to use three elementary metrics.

Let us consider these methods in more details. Below we denote them as *Median(3)*, *Median(5)*, *Alpha-trim(5)*, *Multiplied(2)*, *Multiplied(3)*, respectively. *Median(3)* and *Median(5)* perform as

$$M3 = median(M_1^{lin}, M_2^{lin}, M_3^{lin}) \tag{2}$$

$$M5 = median(M_1^{lin}, M_2^{lin}, M_3^{lin}, M_4^{lin}, M_5^{lin}) \tag{3}$$

$$Alpha-trim(5) = \sum_{q=2}^{4} M_{(q)}^{lin} / 3 \tag{4}$$

where M_i^{lin} is *i*-th elementary no-reference metric after fitting and linearization, $M_{(q)}^{lin}$ denotes the *q*-th order statistic. Thus, *Alpha-trim(5)* performs trimming of the largest and smallest values with averaging the remaining ones.

To obtain the resulting metric, we have made a fitting of metrics for each of the test image databases that have a set of MOS values for each image. Linearization was carried out taking into account the results of the analysis and the recommendations in [6]. For the linearization, the function "Power2" was chosen since it provides the best stability and accuracy. The correlation indices of the combined metric for each of test images database are given in Table 2. The best combination for each combination method from all available ones has been determined by finding the maximum value of SROCC.

Recall that the robust methods of combining elementary metrics, such as *Median(3)*, *Median(5)* and *Alpha-trim(5)* have been successful applied in design of full-reference quality metrics [6] for which SROCC values exceeds 0.8. After applying them to no-reference metrics, the robust methods do not lead to significant positive results. For all test image databases, except MDID and KonIQ10k, their performance does not improve significantly compared to the best elementary metric (see data in Table 2).

There is practically no improvement for the database LIVE (although SROCC values for this database are already high enough). There is no improvement (compared to the best elementary metric, although note that the best elementary metrics are different for different databases) for the databases LIVE MD, NRTID, and TID2013. For the last two databases, MDID and konIQ10k, the SROCC increase is more noticeable reaching 0.06 and 0.05, respectively, for the best combinations.

The metrics *Multiplied(2)* and *Multiplied(3)* after optimization of parameters provide larger improvement of SROCC. It is small for LIVE and LIVE MD databases, reaching 0.03 for NRTID, TID2013 and KonIQ10k, and about 0.06 for MDID. Such results are achieved using different combinations of metrics (individual for each database) and this can cause problems in practice.

Table 2 SROCC values of simple combined metrics

Method	SROCC	Names	Parameters
LIVE			
Single IQA	0.9582	CDIIVINE	
Median(3)	0.9545	BLIINDS2, BRISQUE, CDIIVINE	
Median(5)	0.9525	BLIINDS2, BRISQUE, CDIIVINE, DIIVINE, NJQA	
Alpha-trim(5)	0.9567	BLIINDS2, BRISQUE, CDIIVINE, FISH, NIQE	
Multiplied(2)	0.9624	BLIINDS2, CDIIVINE	0.20, 0.70
Multiplied(3)	**0.9624**	BLIINDS2, CDIIVINE, NIQE	0.25, 0.75, 1.0
LIVEMD			
Single IQA	0.8769	ILNIQE	
Median(3)	0.8637	DIIVINE, ILNIQE, NIQE,	
Median(5)	0.8507	BLIINDS2, DIIVINE, ILNIQE, NIQE, SMETRIC,	
Alpha-trim(5)	0.8737	BLIINDS2, CDIIVINE, ILNIQE, NIQE, NJQA	
Multiplied(2)	0.8859	BLIINDS2, ILNIQE	0.30, 1.00
Multiplied(3)	**0.8863**	BLIINDS2, CDIIVINE, ILNIQE	0.25, 0.25, 1.00
NRTID			
Single IQA	0.7098	SMETRIC	
Median(3)	0.7020	CDIIVINE, DESIQUE, SMETRIC	
Median(5)	0.7017	BRISQUE, CDIIVINE, FISH, NJQA, SMETRIC	
Alpha-trim(5)	0.7131	BRISQUE, CDIIVINE, FISH, NJQA, SMETRIC	
Multiplied(2)	0.7398	CDIIVINE, SMETRIC	0.70, 1.00
Multiplied(3)	**0.7424**	CDIIVINE, DESIQUE, SMETRIC	0.75, 0.25, 1.0
TID2013			
Single IQA	0.4921	ILNIQE	
Median(3)	0.4817	DESIQUE, DIIVINE, ILNIQE	
Median(5)	0.4729	BRISQUE, DESIQUE, DIIVINE, ILNIQE, SMETRIC	
Alpha-trim(5)	0.4831	BIQI, CDIIVINE, DESIQUE, ILNIQE, NJQA	
Multiplied(2)	0.4938	DIIVINE, ILNIQE	0.50, 0.70
Multiplied(3)	**0.5260**	ILNIQE, NIQE, SMETRIC	1.75, -0.5, -1.25
MDID			
Single IQA	0.6942	ILNIQE	
Median(3)	0.7212	BIQI, BRISQUE, ILNIQE	
Median(5)	0.7324	CDIIVINE, ILNIQE, NIQE, NJQA, SMETRIC	
Alpha-trim(5)	0.7432	BIQI, BRISQUE, ILNIQE, NIQE, NJQA	
Multiplied(2)	0.7518	BIQI, BRISQUE	-0.10, -0.10
Multiplied(3)	**0.7617**	BIQI, BRISQUE, SMETRIC	-0.50, -0.50, -0.25
KONIQ10K			
Single IQA	0.6130	SMETRIC	
Median(3)	0.6330	CDIIVINE, FISH, SMETRIC	
Median(5)	0.6472	CDIIVINE,FISH, ILNIQE, NJQA, SMETRIC	
Alpha-trim(5)	**0.6514**	CDIIVINE,FISH, ILNIQE, NJQA, SMETRIC	
Multiplied(2)	0.6369	BIQI, FISH	-0.3, -0.4
Multiplied(3)	0.6442	CDIIVINE, NIQE, SMETRIC	-0.25, -0.25, -0.50

At the same time, the above results do not allow to conclude how effective is the use of elementary metrics and whether it is possible to select a combination of them which will provide stable results for different image databases. Table 3 presents some useful statistics that show how many times each elementary metric hits in the top 5 for different image sets. Among these metrics, it is worth marking BIQI, CDIIVINE, and ILNIQE which are used the most often. The remaining columns demonstrate how often these metrics are used by each of the best combined ones.

Table 3 Statistics of the metrics use in the combined methods

Metric	Single (in TOP5 by each database)	Median (3)	Median (5)	Alpha-trim(5)	Multiplied (2)	Multiplied (3)
BIQI	5	1	0	2	2	1
BLIINDS2	2	1	2	2	2	2
BRISQUE	3	2	3	3	1	1
CDIIVINE	5	3	4	5	2	4
DESIQUE	0	2	1	1	0	1
DIIVINE	3	2	3	0	1	0
FISH	2	1	2	3	1	0
ILNIQE	5	3	4	4	2	2
NIQE	3	1	2	3	0	3
NJQA	0	0	4	5	0	0
SMETRIC	2	2	5	2	1	4

Considering the obtained data, we analyze the possibility of creating a universal combined metric. The disadvantage of the *Multiplied(2)* and *Multiplied(3)* methods is the necessity to optimize their parameters for each image database (this necessity is not so actual for robust methods of metrics' combining). Among combined metrics, *Alpha-trim(5)* provides the highest correlation rates, thus, this method can be used. The metrics CDIIVINE, ILNIQE, and NJQA were used the most times in different image sets. The first two separately provide the maximum result in out of six databases. For this reason, it is worth including SMETRIC, which provides maxima for NRTID and Koniq10k. Among metrics BRISQUE, FISH and NIQE, which were used three times, the first provides more stable results in Table 1, therefore, it will be considered as well. The results of the method of combining *Alpha-trim(5)* with the metrics of BRISQUE, CDIIVINE, ILNIQE, NJQA and SMETRIC are summarized in Table 4. It also shows the position of this metric compared to the elementary metrics in Table 1.

Table 4. Combined metric results for different image databases

	TID 2013	NRTID	LIVE	LIVE MD	MDID	Koniq 10k
Max single IQA	0.4921	0.7098	0.9582	0.8769	0.6942	0.6130
Alpha-trim	0.4688	0.6610	0.9286	0.7548	0.6584	0.6150
Position compared to the best particular metrics						
	#2	#2	#3	#3	#2	#1

As can be seen from Table 4, *Alpha-trim(5)* with BRISQUE, CDIIVINE, ILNIQE, NJQA and SMETRIC metrics without showing maximum results stably works for all image databases, providing high reliability.

Configuration and parameters of no-reference combined neural network metric

Another way to combine elementary metrics is to use learning-based methods: neural networks [33-34], support vector machines [24], data clustering approaches [35], and others. Our work below addresses a possibility of using neural networks for the aforementioned purpose.

Neural networks can be an effective instrument to solve such problems [6]. For this, it is necessary to satisfy a number of requirements imposed on the network during its training. This stage is crucial. The key factors include the quality of the training set of test images, the training sample, the number of incoming data, the network type, the number of layers and neurons in each of them, activation functions, and other parameters. It should be also taken into account that a trained neural network (NN) will not work on data that sufficiently outgo from the training sample.

The development of a combined metric based on an artificial neural network assumes that the values of elementary metrics are fed to the input, and the learning goal is the MOS values, the maximum of which should be achieved by the network during the learning process. The learning phase is a multi-parameter nonlinear optimization resulted by matrices of weights for all incoming metrics. A combined neural network metric (CNNM) will be an aggregation of all these interconnected matrices for each metrics and neuron layers.

As previously defined in [6], preparation of the neural network training stage requires compliance with the following requirements:

1) Input data must be independent, which is ensured by the use of metrics with different mathematical core and properties;

2) Representative training set; in accordance with this requirement, the training set should contain the results of metrics for images with maximal possible number of distortion types with the largest range of their variation.

3) Selection of the neural network configuration (network type, number of layers and neurons, etc.).

In accordance with the specified requirements, let us analyze the data used and the measures taken to prepare the training sets:

1) The metrics differ significantly, which is confirmed by the results in Table 1. Even ILNIQE and NIQE metrics have sufficient difference in their performance.

2) To ensure maximum representativeness of the data, each set of test images was divided into two subsets (training and test). Image indices were divided using a random distribution function according to a uniform law. To improve the accuracy of training, which directly depends on the amount of input data (some of which will be used not for training, but for validation) for each image database for the training set, 70% of images from and 30% for testing are allocated. In order to minimize possible negative factors of distribution unevenness, three such distributions were formed for each image database.

3) Configuration of neural networks. To solve the third task, we have to choose a good NN. There are no strict recommendations on choosing the best ones. Thus, we have used several types and configurations of NN as well as varied their parameters. Based on this analysis, it has become possible to choose the best configuration.

The neural networks of the following types have been used:

- Feed forward back propagation network, 'feed' (fig.1a);
- Cascade forward back propagation network, 'cascade' (fig.1b);
- Elman back propagation network, 'elman' (fig.1c);
- Generalized regression network, 'genreg' (fig.1d);
- Layer recurrent network (similarly to Elman network, fig.1c); NARX (Nonlinear autoregressive neural network) on fig.1f.

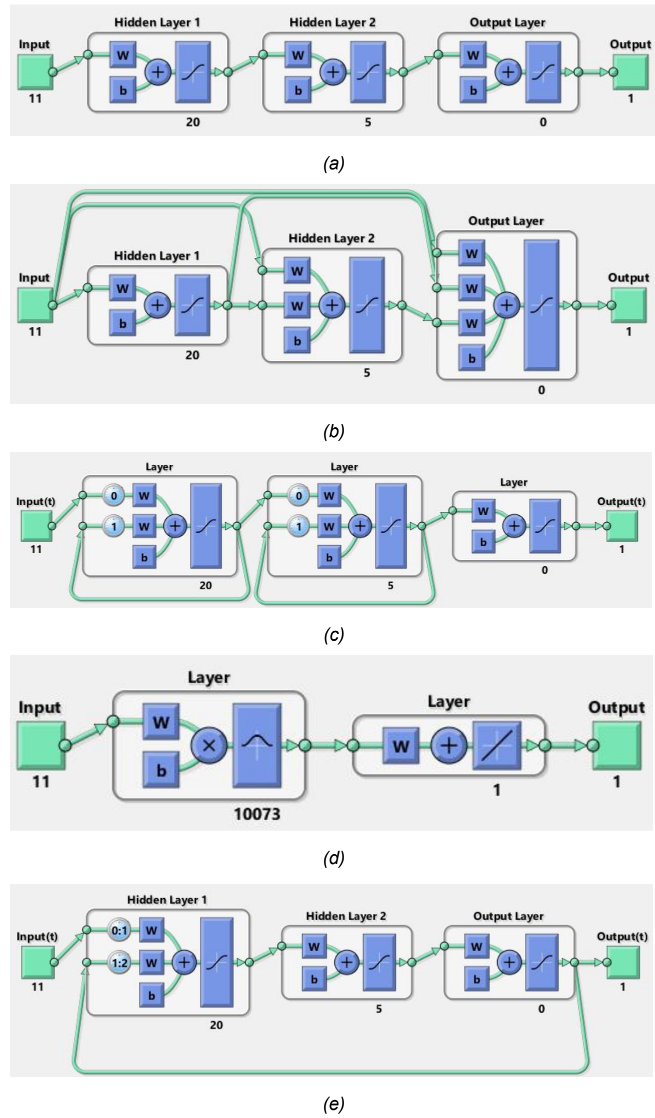

Figure 1. Neural networks schemes.

For each of the considered types of networks, the following ten configurations of layers and neurons were created:

- one hidden layer: [5], [20].
- two hidden layers: [5, 5], [20, 20], [5, 20], [20, 5],
- three hidden layers: [5,10,20], [5,5,5], [20,20,20], [20,10, 5].

Neural network training is the process of finding a global minimum of the root-mean-square error with a random initial bias in accordance with a given adaptation function of the training function. Therefore, to eliminate the problem of non-correct learning, when the function stops at the local minimum, training of minimum 20 networks was started for each configuration. In total, these are over 21,000 configurations of various trained networks.

Neural networks optimization and analysis

Table 5 contains data only for two best NN-based metrics that provide the best performance according to SROCC for test subsets of each database. Column "#" shows the number of image distribution for given database. After analysis of the obtained neural networks we should note, that it is important factor for networks learning and further works. For the best metrics by different image subsets the deviations of the SROCC values reached a range 0.05...0.06.

Table 5. The best CNNM by SROCC value of test image subsets for all databases.

Metrics by dataset (rank)	#	Network type	Layers configuration	SROCC (train)	SROCC (test)	Single IQA
Live (1)	1	elman	[20 5]	0,9823	0,9822	0.9582
Live (2)	3	layrec	[5 20]	0,9795	0,9824	0.9582
Livemd (1)	1	cascade	[5 5 5]	0,8977	0,9083	0.8769
Livemd (2)	1	elman	[20 20]	0,9107	0,9069	0.8769
Tid2013 (1)	3	cascade	[20 20 20]	0,7763	0,7720	0.4921
Tid2013 (2)	3	elman	[20 20 20]	0,7655	0,7616	0.4921
Nrtid (1)	1	elman	[20 10 5]	0,7771	0,8040	0.7098
Nrtid (2)	1	layrec	[20 20 20]	0,7696	0,7995	0.7098
Mdid (1)	2	layrec	[5 5 5]	0,8551	0,8531	0.6942
Mdid (2)	2	cascade	20	0,8829	0,8504	0.6942
Koniq10k (1)	2	elman	[20 5]	0.7030	0.6911	0.613
Koniq10k (2)	3	cascade	[5 10 20]	0.7063	0.6910	0.613

According to the results in Table 5, we would like to note several aspects. The combined NN-based metrics on the test set provide sufficiently better performance compared to simple combination methods considered in the previous Sections. For example, for TID2013, the multiplied metric has shown maximum SROCC=0.5260, at that time several neural networks have SROCC>0.76. It is also worth stressing that the results for training and test sets are very close.

Table 6 Results of combined metrics' verification for each image database

Metrics by dataset (rank)	LIVE	LIVE MD	TID 2013	NRTID	MDID	KonIQ 10K
Live (1)	**0.982**	0.486	0.193	*0.663*	0.417	0.277
Live (2)	**0.982**	0.557	0.249	*0.728*	0.061	0.379
Livemd (1)	0.800	**0.908**	0.361	0.120	0.411	0.347
Livemd (2)	0.090	**0.907**	0.097	0.546	0.280	*0.576*
Tid2013 (1)	0.706	0.116	**0.772**	0.570	0.153	0.160
Tid2013 (2)	0.365	*0.796*	**0.762**	0.224	0.539	0.403
Nrtid (1)	0.215	0.688	0.401	**0.804**	0.399	0.350
Nrtid (2)	0.053	0.530	0.381	**0.800**	*0.613*	0.064
Mdid (1)	0.060	0.127	0.377	0.399	**0.853**	0.512
Mdid (2)	0.810	0.693	0.252	0.457	**0.850**	0.079
Koniq10k(1)	0.481	0.019	0.063	0.440	0.258	**0.691**
Koniq10k(2)	***0.924***	***0.777***	***0.470***	0.604	***0.713***	**0.691**
Single IQA	0.958	0.877	0.492	0.710	0.694	0.613
Alpha-trim	0.929	0.755	0.469	0.661	0.658	0.615
Cascade (20,5), 1set	0,882	*0,833*	***0,676***	*0,722*	*0,702*	*0,628*
NARX (5), 3 set	*0,940*	***0,870***	*0,463*	*0,719*	*0,717*	*0,624*

Comparing the networks' types, the best performance was often provided by neural networks with feedback ('elman' and 'layrec'). Mostly they use maximum possible number of layers and neurons. They demonstrate the most accurate adaptation to the image test set. However, at the same time, such NN-based metrics occur to be adapted to a particular image database and they can perform differently for others. This fact is shown by the results presented in Table 6. In the leftmost column, two best NN-based metrics from Table 5, are presented. In other columns, SROCC values are presented for each metric under condition that the metric is applied to the database it was trained (diagonal marked by Bold) and other databases.

One can see that, trained on images of a particular database, the NN-based metrics occur less suitable for others. Some of them show acceptable results for 2 or 3 databases, but their overall effectiveness is low. The only exception was the trained metric denoted as *Koniq10k(2)*. Although it does not demonstrate too high performance for the databases LIVEMD and NRTID, the metric *Koniq10k(2)* has no obvious failures. It outperforms robust metrics (the results for one of them is given in Table 6) and is often better than the best elementary metrics individual for each database.

Universality of other NN-based metrics not mentioned in Table 5 has been analyzed. Two good ones (according to average SROCC values) are presented in Table 6 in the lowest rows. Their performance is comparable to Koniq10k(2).

Although a question of training is not fully resolved yet, we can state that the NN-based metrics provide better results than other versions of the combined metrics. However, the problem of metric's universality still remains.

Conclusions

This paper deals with the problem of low accuracy of no-reference image visual quality metrics and their poor operation on certain sets of test images. Creation of simple combined metrics based on robust estimates is considered. Fitting and linearization aspects also have been discussed.

In order to further improve the performance of no-reference assessment of visual quality, a combination method based on neural networks was considered. Several types of such networks with different configurations and the influence of various factors on the final accuracy of such metrics have been studied. Several good metrics that outperform elementary ones and work well for different databases, have been designed.

REFERENCES

[1] Keelan, B.W., Handbook of Image Quality, Marcel Dekker Inc., 2002.

[2] R. C. Gonzalez, R. Woods, Digital Image Processing, Pearson, 2018.

[3] W. Lin and C. C. Jay Kuo, "Perceptual visual quality metrics: A survey," Journal of Visual Communication and Image Representation, vol. 22, no. 4, pp. 297-312, 2011.

[4] D. M. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present, and Future Research," ISRN Signal Processing, vol. 2013, pp. 1-53, 2013.

[5] A. K. Moorthy and A. C. Bovik, "Visual Quality Assessment Algorithms: What Does the Future Hold?," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 675-696, 2011.

[6] O. Ieremeiev, V. Lukin, N. Ponomarenko, K. Egiazarian, "Robust linearized combined metrics of image visual quality," Electronic Imaging, Image Proc.: Algorithms and Systems XVI, pp. 1-6, 2018.

[7] V. V. Lukin, N. N. Ponomarenko, O. I. Ieremeiev, K. O. Egiazarian and J. Astola, "Combining of full-reference image visual quality metrics by neural network," in Proceedings of SPIE 9394 Human Vision and Electronic Imaging XX, San Francisco, 2015.

[8] L. Jin, K. Egiazarian and C. C. Jay Kuo, "Perceptual image quality assessment using block-based multi-metric fusion BMMF," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, 2012.

[9] Z. Wang, A. Bovik, "Reduced- and no-reference image quality assessment," IEEE Sign. Process. Mag., vol. 28, no 6, pp.29-40,2011.

[10] Okarma, "Colour Image Quality Assessment Using the Combined Full-reference Metric," Computer recognition Systems 4, Advances in Intelligent and Soft Computing, vol. 95, pp. 287-296, 2011.

[11] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1098–1105, 2012.

[12] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," IEEE Transactions on Image Processing, vol. 27(1), pp. 206-219, 2018.

[13] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics, Advances of Modern Radioelectronics, 10 (2009) 30-45, TID2008 page: http://ponomarenko.info/tid2008.htm, 7 Nov. 2018.

[14] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battistid, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," Journal of Signal Processing: Image Communication, vol. 30, pp. 57–77, 2015.

[15] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE Image Quality Assessment Database Release 2, in http://live.ece.utexas.edu/research/quality/subjective.htm, 12 Nov. 2018.

[16] H. Lin, V. Hosu, D. Saupe, "KonIQ-10K: Towards an ecologically valid and large-scale database" in http://database.mmsp-kn.de/koniq-10k-database.html, 27 Sep. 2018.

[17] N. Ponomarenko, O. Eremeev, V. Lukin, K. E., "Statistical evaluation of no-reference image visual quality metrics," in 2010 2nd European Workshop on Visual Information Processing (EUVIP), Paris, 2010.

[18] D. Jayaraman, A. Mittal, A. K. Moorthy and A. C. Bovik, "Objective Quality Assessment of Multiply Distorted Images," in Proceedings of Asilomar Conference on Signals, Systems and Computers, 2012.

[19] W. Sun, F. Zhou, Q. M. Liao. "MDID: a multiply distorted image database for image quality assessment," Pattern Recognit. 61C, pp. 153-168, 2017.

[20] O. Ieremeiev, N. Ponomarenko, V. Lukin, "Evaluation of accuracy characteristics of the method of visual quality metrics verification based on specialized image databases (in Russian)," Radio and electronic systems, vol 61, pp. 48-57, 2013.

[21] A. K. Moorthy and A. C. Bovik, ―A two-step framework for constructing blind image quality indices,‖ IEEE Signal Processing Letters, vol. 17,no. 2, pp. 587–599, May 2010.

[22] M. A. Saad, A. C. Bovik and C. Charrier, `` Model-Based Blind Image Quality Assessment: A natural scene statistics approach in the DCT domain' ", IEEE Transactions Image Processing, pp. 3339-3352, vol. 21, no. 8, 2012.

[23] A. Mittal, A. K. Moorthy and A. C. Bovik, "No Reference Image Quality Assessment in the Spatial Domain", IEEE Transactions on Image Processing, vol. 21, issue 12, pp.4695 - 4708, 2012.

[24] Y. Zhang, A. K. Moorthy, D. M. Chandler, and A. C. Bovik, "CDIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes," Signal Processing: Image Communication, vol. 29, no. 7, pp. 725-747, 2014.

[25] Y. Zhang, D. Chandler, "An Algorithm for No-Reference Image Quality Assessment Based on Log-Derivative Statistics of Natural Scenes," in Proceedings of SPIE - The International Society for Optical Engineering 22(4), 2013

[26] A. K. Moorthy and A. C. Bovik, `` Blind Image Quality Assessment: From Scene Statistics to Perceptual Quality '', IEEE Transactions Image Processing, vol. 20, no. 12, pp. 3350-3364, 2011.

[27] P.V. Vu, D.M. Chandler , "A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation," IEEE Signal Processing Letters, vol.19, no.7, pp.423-426, 2012.

[28] Lin Zhang, Lei Zhang, and A.C. Bovik, "A feature-enriched completely blind image quality evaluator," IEEE Transactions on Image Processing, vol. 24, Issue 8, pp. 2579 - 2591, 2015.

[29] A. Mittal, R. Soundararajan and A. C. Bovik, " Making a Completely Blind Image Quality Analyzer ", IEEE Signal processing Letters, vol. 22, no. 3, pp. 209-212, 2013.

[30] S. A. Golestaneh, D. M. Chandler, "No-Reference Quality Assessment of JPEG Images via a Quality Relevance Map," IEEE Signal Processing Letters, vol.21, pp. 155-158, 2014.

[31] N. N. Ponomarenko, V. V. Lukin, O. I. Eremeev, K. O. Egiazarian, J. T Astola, "Sharpness metric for no-reference image visual quality assessment," in Proceedings of SPIE 8295, Image Processing: Algorithms and Systems X

[32] O. I. Ieremeiev, V. V. Lukin, N. N. Ponomarenko, K. O. Egiazarian, J. Astola, "Combined full-reference image visual quality metrics," Electronic Imaging, Image Proc.: Algorithms and Systems XIV, vol.15, pp.1-10, 2016.

[33] Bishop, C. M., "Pattern Recognition and Machine Learning (Information Science and Statistics)", Springer, New York, 2006.

[34] Mitchell, T.M., "Machine Learning," McGraw Hill, New York, 1997..

[35] R.Xu, D. C. Wunsch, "Clustering", Wiley-IEEE Press, New Jersey, 368p. (2008).