Image Quality Assessment Using Reduced-Reference Nonlinear Model

Dragana Dørđević*, Dragan Kukolj**, Maja Pokrić*, Ivana Ostojić*,
* RT-RK Institute for Computer Based Systems, Narodnog fronta 23a, Novi Sad, Serbia
** University of Novi Sad, Faculty of Technical Sciences, Serbia
dragana.djordjevic@rt-rk.com, dragan.kukolj@rt-rk.com, maja.pokric@rt-rk.com, ivana.ostojic@rt-rk.com

Abstract— This paper presents a reduced-reference nonlinear model driven image quality scheme that is based on a Neural Network statistical estimator, namely Multilayer Perceptrons (MLP) and that is optimized to the Mean Opinion Score (MOS) scale for the combination of input different objective quality measures. In order to examine the performance of the models and identification of how well the model estimates the MOS as a reference model, a linear model is chosen. Also, in this paper is analyzed the effects of an input measures selection on the quality of the MOS estimation.

Keywords— Reduced-reference Model, Image Quality Assessment, Mean Opinion Score

I. INTRODUCTION

There is an increased need to measure and assess the quality of images/video sequences. The perceived quality greatly depends on the image/video compression codec and the employed bit rate, but also on the content.

The objective assessment of image quality is well-researched topic. The problem of image quality assessment in multimedia has been addressed in many papers and standards [1][2]. The most objective quality measures try to model the behavior of the Human Visual System (HVS), which is very complex and currently not completely understood. In order to improve the prediction performance, a new type of quality measure has been proposed that is based on Machine Learning (ML). In this approach, a ML system is trained on a subjective quality assessment database and applied on a set of perceptually relevant input objective quality measures.

The main principle is to find the best combination of the features, extracted from the pre-defined training image with different levels of degradation. As a result, a ML-based quality scheme tries to mimic the behavior of the HVS to predict the perceived image quality, rather than explicitly modeling it. Hence, its prediction performance does not depend on our knowledge of the HVS.

In recent years, a considerable number of papers have addressed the problem of combining different objective measures/features to design a reliable system for overall perceptual image quality assessment. While some authors employ relatively simple polynomial models to combine the features [3] [4] (most often in the form of a weighted Minkowski metric [5]), others propose to use more advanced learning schemes, such as Artificial Neural Network (ANN) [6-11].

A variety of ANN’s types and configuration with aim to estimate perceived quality of images or video content is proposed. A methodology for MPEG-2 video streams quality estimation using circular back-propagation (CBP) neural networks is presented in [11]. An objective features continuously extracted from frames of MPEG-2 video streams are fed into the CBP ANN estimating the corresponding perceived quality.

On the other hand, Fiegel proposed the Radial Basis Function Network (RBFN) in [6] and compares it with previously used Multi Layer Perceptron (MLP). The inputs of the NN are extracted from the original and the distorted images and decomposed into several bands. He concluded in his work that the performance of MLP model can be improved while utilizing data preprocessing.

Further, Kukolj et al. in [8] proposed a novel scheme for selecting the most relevant features from large set of inputs and used that selected features for the design of Modular Neural Network (MNN) structure scheme. The proposed model is also based on the reduced features space clustering which enables the training of separate models related to the corresponding video content and the actual MOS.

In the literature, it should be noted that is a very common usage of MLP network configuration. This occurrence is due to the fact that this configuration provides very effective results in mapping the data in the field of image quality assessment. Thus, in [7] the authors used full-reference feed-forward MLP approach while no-reference feed-forward MLP approach is used in [9] and [10].

For this purpose, a reduced-reference nonlinear model that is optimized to the Mean Opinion Score (MOS) scale for the combination of different objective quality measures or features is developed. The reduced-reference model is designed to predict the perceptual quality of distorted image by using certain features of the original signal. The framework model is computed only once in the training phase and is based on a set of image artifacts, represented through selected features.

The remainder of this paper is structured as follows. In Section 2, an overview of the selected input objective quality measures is given, while in Section 3 these input quality measures are combined by the nonlinear model that is specifically adopted for the purpose of the image quality assessment. The experimental results and comparison with reference linear model are given in Section 4 along with the explanations of the nonlinear model performances and characteristics. The conclusions are drawn in Section 5.
II. SELECTION OF INPUT QUALITY MEASURES

Mathematical models that automatically approximate results of subjective quality assessment, called objective image techniques, are based on criteria and metrics that can be measured objectively. Objective models which approximate scores given by human observers could be classified in three categories: no-reference, reduced-reference and full-reference model. No-reference metrics consumes only information from the distorted image and based on this information could produce quality score for image. Reduced-reference metrics requires both information, from original and distorted image and compare them to given quality score. On the other hand, full-reference metrics are ideal solution but rarely applicable in practice because it is unlikely that there is all information about the original image. Full-reference metrics compare all information from distorted image to original image.

In this paper four objective quality measures, labeled by $d_{SI}$, $d_{CON}$, $d_{HIST}$ and $d_{JPEG}$ are considered. The selected quality measures, described in [13], are specifically designed to measure a specific aspect of perceptual quality.

The $d_{SI}$ quality metric, namely structural information index was proposed initially in [14]. This feature is based on statistic and compares only edge information [15] of original undistorted image and degraded image. This comparison detects spatial impairments, such as blurring and blocking, more precisely structural information loss.

The quality metric $d_{CON}$ is one of three components of Structural Similarity Index (SSIM) [16] and it is highly effective to detect spatial noise. The contrast quality measure is locally calculated on certain number of image blocks. The final output is obtained by averaging the local scores.

The $d_{HIST}$ quality measure can be used for the quality assessment of JPEG compressed images. The reference and distorted image are divided into 25 equal-sized blocks $x_i$ and $y_i$ with $i \in \{1,2,...,25\}$. Let $H_{x_i}$ and $H_{y_i}$ denote the corresponding block histograms and determine the local quality scores. The local quality score is low in case of detail attenuation and high in case of compression artifacts. As JPEG compression simultaneously causes detail attenuations in some blocks and compression artifacts like ringing and blocking in other blocks, the variability in local scores will be high.

The last metric $d_{JPEG}$ is a no-reference quality measurement algorithm for JPEG compression developed in [3]. The quality estimation is based on three image features that measure the intensity of the blocking artifacts and the attenuation of detail in the DCT blocks.

III. QOE USING NONLINEAR MODEL WITH REDUCED-REFERENCE QUALITY METRICS

The proposed image quality assessment approach is a reduced-reference nonlinear model driven image quality prediction scheme. Nonlinear quality model based on ANN prediction is created to optimally predict the perceptual quality, expressed in the Mean Opinion Score (MOS) scale. As a reference model in the paper the reduced reference linear model is used, precisely linear regression. In the proposed image quality assessment modeling scheme several stages can be distinguished.

The influence of the input features to the system was examined firstly and how those input features match the actual MOS by using the Pearson and Spearman correlation. Table 1 summarizes these correlations.

<table>
<thead>
<tr>
<th>Input features</th>
<th>Pearson</th>
<th>Spearman</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{CON}$</td>
<td>0.810</td>
<td>0.888</td>
</tr>
<tr>
<td>$d_{HIST}$</td>
<td>0.554</td>
<td>0.632</td>
</tr>
<tr>
<td>$d_{SI}$</td>
<td>0.593</td>
<td>0.582</td>
</tr>
<tr>
<td>$d_{JPEG}$</td>
<td>0.177</td>
<td>0.003</td>
</tr>
</tbody>
</table>

From Table 1 it can be conclude that the input feature $d_{CON}$ will have the greatest impact on training because the both correlation coefficients are close to value 1. The next two input features $d_{HIST}$ and $d_{SI}$ indicate partial correlation and can be expected that involvement in the training of the neural network could realize improvements, while the poor correlation of $d_{JPEG}$ will not provide significant improvements.

The second step was setting a MLP neural network. The MLP is configured to contain a single hidden layer, maximum four input neurons which is equal to number of input features and hidden nodes formulated by $2n+1$ while $n$ is number of input features as shown in Figure 1. For activation function a tangent-hyperbolic for hidden nodes was chosen while linear function was chosen for output node. The MLP was trained with Levenberg-Marquardt optimization back-propagation algorithm.

<table>
<thead>
<tr>
<th>Activation function</th>
<th>Hidden layer</th>
<th>Output layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>tangent-hyperbolic</td>
<td>dcon</td>
<td>dhist</td>
</tr>
<tr>
<td></td>
<td>dsi</td>
<td>djpeg</td>
</tr>
</tbody>
</table>

Figure 1. MLP configuration

IV. RESULTS

In those experiments a set of 24 reference images obtained from public available image database TID2008 [17] were used. Image degradation was performed by using three types of distortion "Gaussian blur", "JPEG compression" and "JPEG2000 compression" with four levels of degradation, and therefore the total number of distorted images was 24x3x4 = 288. The proposed nonlinear model on these 288 images with four input
reduced-reference image quality features: \( d_{SI} \), \( d_{HIST} \), \( d_{CON} \) and \( d_{JPEG} \) described in Section 2 was applied. Each of these functions is focused on a specific type of distortions. During the experiments ten-fold cross-validation in order to ensure the exactness of the results was used thus the ratio of training and test images were 90% and 10% of total number of images.

In Table 2 the results of the proposed image quality assessment are shown for different number of input features and models incorporated in terms of root mean square error (RMSE), standard deviation of RMSE, Spearman and Pearson statistical measures.

<table>
<thead>
<tr>
<th></th>
<th>Linear model all input features</th>
<th>Nonlinear model all input features</th>
<th>Nonlinear model ( d_{CON} ) + ( d_{HIST} ) + ( d_{SI} )</th>
<th>Nonlinear model ( d_{CON} ) + ( d_{HIST} )</th>
<th>Nonlinear model ( d_{CON} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>0.909</td>
<td>0.987</td>
<td>0.977</td>
<td>0.953</td>
<td>0.886</td>
</tr>
<tr>
<td>Spearman</td>
<td>0.914</td>
<td>0.980</td>
<td>0.969</td>
<td>0.949</td>
<td>0.894</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.707</td>
<td>0.275</td>
<td>0.351</td>
<td>0.471</td>
<td>0.754</td>
</tr>
<tr>
<td>Standard deviation of RMSE</td>
<td>0.001</td>
<td>0.014</td>
<td>0.012</td>
<td>0.036</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Spearman and Pearson statistical measures were used for testing correlation between estimated MOS, output of predicted model, and the actual MOS scores. It can be seen based on data from Table 2 that both measures show satisfactory level of correlation for all models, even in case of the linear model.

If the RMSE values were compared, it can be observed that the estimated proposed nonlinear model with all input features (RMSE = 0.275) gives a much smaller error than referent linear model (RMSE = 0.707) which is one of the expected results. Looking at the other nonlinear models with less input features (but not with only one input), smaller error was given than the linear model and the conclusion is that neural networks is a good choice for predicting subjective images score. On the other hand, it can be noticed that the number of input features drastically affect the estimation process and it can be seen by reducing the number of input parameters increases the error. For example, error for 3 input features is RMSE = 0.351 but for only one input feature is greater than linear model (RMSE = 0.754).

Comparing MOS obtained from subjective evaluation and estimated MOS, as shown in Figure 2, it can be concluded that all models are more accurate for the samples with less degradation. Also, it can be noticed that nonlinear model in Figure 2 (b) has very good results in the form of the estimated MOS that perfectly mapped the true MOS. By reducing the number of inputs features, prediction of MOS significantly deteriorates as can be seen by comparing the graphic (b) and (c) of Figure 2. Figure 3 shows the distribution of points depending on the level of distortion in two-dimensional space of input parameters \( d_{HIST} \) and \( d_{CON} \) (comparison contrast component of the SSIM [16]). The points in Figure 3 with higher levels of distortion are grouped towards the lower left corner, while other with a smaller level of distortion group towards the top right corner of the chart.
of objective metrics as its inputs. The improvement of the model performances is fully compatible with correlations between an objective metrics and an image subjective rating.

ACKNOWLEDGMENT

This work was partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia under Grant TR-32034 and by the Secretary of Science and Technology Development of the Province of Vojvodina under Grant 114-451-2636/2012-01.

REFERENCES