GAP-RBF Based NR Image Quality Measurement for JPEG Coded Images

R. Venkatesh Babu[1] and S. Suresh[2]

[1] Department of Electrical Engineering
Indian Institute of Science, Bangalore, India
venkatesh.babu@gmail.com
[2] School of EEE,
Nanyang Technological University, Singapore
suresh99@gmail.com

Abstract. In this paper, we present a growing and pruning radial basis function based no-reference (NR) image quality model for JPEG-coded images. The quality of the images are estimated without referring to their original images. The features for predicting the perceived image quality are extracted by considering key human visual sensitivity factors such as edge amplitude, edge length, background activity and background luminance. Image quality estimation involves computation of functional relationship between HVS features and subjective test scores. Here, the problem of quality estimation is transformed to a function approximation problem and solved using GAP-RBF network. GAP-RBF network uses sequential learning algorithm to approximate the functional relationship. The computational complexity and memory requirement are less in GAP-RBF algorithm compared to other batch learning algorithms. Also, the GAP-RBF algorithm finds a compact image quality model and does not require retraining when the new image samples are presented. Experimental results prove that the GAP-RBF image quality model does emulate the mean opinion score (MOS). The subjective test results of the proposed metric are compared with JPEG no-reference image quality index as well as full-reference structural similarity image quality index and it is observed to outperform both.

1 Introduction

The main objective of image/video quality assessment metrics is to provide an automatic and efficient system to evaluate visual quality. It is imperative that these measures exhibit good correlation with perception by the human visual system (HVS). The most widely used objective image quality metrics, namely mean square error (MSE) and peak signal to noise ratio (PSNR), as widely observed do not correlate well with human perception [1] besides requiring the original reference image to compute distortion. Most images on the Internet and in multimedia databases are only available in compressed form, and hence inaccessibility of the original reference image, makes it difficult to measure the

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 718–727, 2006.

image quality. Therefore, there is an unquestionable need to develop metrics that closely correlate with human perception without needing the reference image.

Considerable volume of research has gone into developing objective image/video quality metrics that incorporate perceived quality measurement with due consideration for HVS characteristics. However, most of the proposed metrics based on HVS characteristics require the original image as reference [2,3,4,5]. Though it is easy to assess the image quality without any reference by manual observations, developing a no-reference (NR) quality metric is a difficult task. To develop NR metrics, it is essential to have *apriori* knowledge about the nature of artifacts. Currently, NR quality metrics are the subject of considerable attention by the research community, visibly so, with the emergence of video quality experts group (VQEG) [6], which is in the process of standardizing NR and reduced-reference (RR) video quality assessment methods.

In recent years, neural networks have emerged as powerful mathematical tools for solving problems as diverse as pattern classification/recognition, medical imaging, speech recognition etc. The increasing popularity of neural networks is due to their ability to construct good approximation of functional relationship between the known set of input and output data. In neural networks, the choice of learning algorithm, number of hidden neurons and weight initialization are important factors in the learning performance. In particular, the choice of learning algorithm determines the rate of convergence, computational cost, and the optimality of the solution. The choice of number of hidden neurons determines the learning and generalization ability of the network. Another important problem is that the re-training process involved in the architectures, whenever we receive a new set of observations (images). Sometimes, the new set of observations may change complexity of the input-output relationship (complexity of the model) and in-turn affects the approximation ability of the neural network model. The process of developing the new neural model with the current training set may leads to increase in computational time. Sequential learning algorithms, which do not require retraining whenever new observation is received, helps to overcome the afore mentioned problems faced by neural network.

In this work, problem of image quality estimation without reference image is reduced to a function approximation problem using GAP-RBF networks. The unknown functional relationship between the HVS features and MOS is captured by the leaning phase of GAP-RBF network. The GAP-RBF quality model is developed with set of 20 source images and its 134 compressed images. The generalization performance of the quality model is evaluated using a new set of 9 source images and its 70 compressed images. The results show that the proposed GAP-RBF model could emulate the MOS effectively compared to the existing techniques.

The paper is organized as follows: Section 2 presents the concepts underlying feature extraction based on various HVS criteria. The basics of GAP-RBFN Image Quality Model are dealt with, in section 3. Subjective test results and discussions are presented in Section 4. Finally Section 5 concludes the paper.

2 HVS-Based Feature Extraction

It is easily deducible that most of the distortion in image/video is due to block DCT-based compression. The most popular and widely used image format, on Internet and digital cameras, happens to be, JPEG [7]. Since JPEG uses block-based DCT transform for coding, to achieve compression, the major artifact that JPEG-compressed images suffer, is blockiness. In JPEG coding, non-overlapping 8×8 pixel blocks are coded independently using DCT transform. The compression (bit-rate) and image quality are mainly determined by the degree of quantization of these DCT coefficients. The undesirable consequences of quantization manifest as blockiness, ringing and blurring artifacts in the JPEG coded image. It turns out that the subjective data for all these artifacts are highly correlated [8]. Hence, measuring the blockiness in-turn indicates the overall image quality.

The proposed NR metric is designed to take into consideration the various human visual criteria while quantifying the blocking artifact. These blocking artifacts would appear as horizontal and vertical edge distortions at the boundaries of 8×8 blocks. The visual sensitivity to these edges is affected by the following parameters [2]: i) Edge Amplitude ii) Edge Length iii) Background Activity and iv) Background Luminance.

The objective of the proposed metric is to integrate the afore-mentioned human visual factors to measure the quality of the JPEG-compressed images. First, we obtain the edges along horizontal and vertical directions using the corresponding 'prewitt' edge operators. Activity along, as well as, on either sides of the horizontal and vertical edges, is captured by high-pass filtering. The final binary activity mask is obtained by hard thresholding the activity measure. This mask only permits regions with lower activity to be considered for blockiness measurements. The background luminance weights are obtained based on the model proposed by Karunasekera et al., [2]. Here darker regions (0 to 127) are given less weight and brighter regions (128 to 255) are given higher weights. Each pixel of the edges that belong to the activity mask is multiplied by the corresponding luminance weight, in order to obtain the obtain final horizontal and vertical edge maps. The horizontal and vertical edge profiles are computed from these weighted edge maps. These profiles indicate the edge strength along each row and column of the weighted edge map. Since the effect of blockiness is seen only at block boundaries, every eighth location of the horizontal and vertical profiles is considered for measuring blockiness. The measure of deviation at every eighth location from the average value of the neighborhood of both (horizontal and vertical) profiles is used for extracting the features. For detailed explanation of feature extraction refer [9].

Since image quality is a subjective phenomenon, the human observer plays a major role in testing image quality metric. The subjective test designates the opinion of a viewer (opinion score) on a given image based on how it is perceived. The mean opinion score (MOS) is the average opinion score over all subjects. The aim of any quality metric is to predict the quality as close as possible to MOS. Hence, the objective here is to find the functional relationship between extracted HSV features and MOS to quantify image quality. In the next section we explain the network architecture and learning algorithm used for quantifying image quality.

3 GAP-RBFN Image Quality Model

In recent years, many sequential learning algorithms have been developed to overcome the problems encountered in neural networks [10,11,12]. Here, radial basis function network (RBF) is used to approximate the functional relationship. These sequential learning algorithms perform better than the batch learning algorithms as they do not require retraining whenever new observations are received. In sequential learning algorithms, the training samples are presented only once to capture the functional relationship whereas in batch learning algorithms, the samples are presented many times. Hence, the sequential learning algorithms require less computational effort and memory requirement than the batch learning algorithms.

Most of the sequential learning algorithms employ some strategy to obtain a compact network to represent input-output relationship. Recently Huang et al., [12] proposed a new sequential learning algorithm called 'growing and pruning radial-basis function (GAP-RBF) network. In this algorithm, the criteria for growing/pruning of hidden neurons is based on the significance of the neurons to the network output. The algorithm updates only the parameters of the nearest neuron to minimize the error. Hence this method is economical from computational as we as memory requirement point of view. Here, we use 'growing and pruning radial-basis function network' to approximate the functional relationship between HVS-based features and MOS. Also, we show that the GAP-RBF based image quality model adapts its features when new image sets are presented. Finally, we compare the performance of the proposed GAP-RBF model with the existing NR and FR image quality metrics [1,13].

The GAP-RBF image quality model is shown in Fig. 1. The basic building block for GAP-RBF is the radial basis function network. In general, a radial-basis function network consists of three layers of processing elements. The first layer linear and only distributes the input signal, while the next layer is nonlinear and uses Gaussian functions. The third layer linearly combines the Gaussian outputs. In a GAP-RBF quality model, the inputs are the extracted HVS features (U) of a given image while the output is the approximated image quality ($\hat{Q}$). The objective is to find a the compact model to approximate the MOS using HVS-based features. The learning algorithm uses 'growing and pruning' strategy to decide on the significance of a neuron towards realizing a compact model. The network parameters such as center vectors, connection weights and widths of hidden neurons are tuned using extended Kalman filter (EKF) algorithm [12].

The output of an GAP-RBF quality model with K Gaussian neurons has the following form:

$$\hat{Q} = \sum_{i=1}^{K} \alpha_i exp\left(-\frac{1}{\sigma_i^2}\|U - \mu_i\|\right) \tag{1}$$

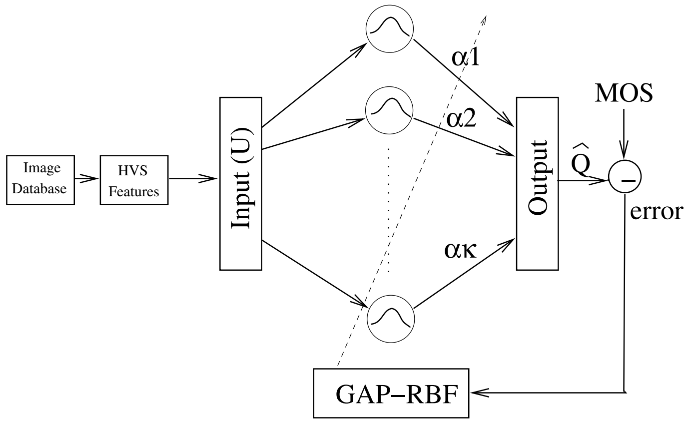


Fig. 1. Overview of the proposed image quality estimation algorithm

where U is the HVS-based feature input vector, α_i is the weight connecting the i^{th} Gaussian neuron to the output neuron, μ_i is the center vector of i^{th} Gaussian neuron and σ_i is the width of the Gaussian neuron.

GAP-RBF is initialized with zero hidden neurons. As new images are received sequentially, the HVS-based features (inputs) are extracted and the network builds up based on a 'growth and pruning' criterion. The algorithm adds/prunes hidden neurons and also tunes the network parameters. Detailed description of the the algorithms can be found in [12]. The following are the steps involved in obtaining a compact network.

3.1 Growing and Pruning Algorithm

Given an approximation error e_{min}, for each observation $(U_n,\ MOS_n)$, where $U_n \in R$, and assuming the number of neurons developed using previous $n-1$ images to be K, the following steps are used to develop the model:

Step 1 **Compute** the estimated image quality for a given image using equation (1)

Step 2 **Growth criterion**: Compute the criterion using the specified parameters

$$\begin{aligned} \epsilon_n &= max\left\{\epsilon_{max}\gamma^n, \epsilon_{min}\right\} \\ e_n &= \hat{Q}_n - Q_n \end{aligned} \tag{2}$$

Step 3 **Adding neuron** based on growth criterion and other conditions
IF $\|U_n - \mu_{nr}\| > \epsilon_n$ and $(1.8.\kappa\|U_n - \mu_{nr}\|)\,|e_n|/S(U) > e_{min}$ then
Allocate $(K+1)$th hidden neuron with

$$\begin{aligned} \alpha_{K+1} &= e_n \\ \mu_{K+1} &= U_n \\ \sigma_{K+1} &= \kappa\|U_n - \mu_{nr}\| \end{aligned} \tag{3}$$

ELSE
Update the network parameters α_{nr}, μ_{nr} and σ_{nr} for the nearest neuron only, using EKF algorithm [12].
Criterion for **pruning** the hidden neurons:
IF $|(1.8\sigma_{nr})^l \alpha_{nr}/S(U)| < e_{min}$, where $S(U)$ is the estimated size of the range where the training samples are drawn from,
remove the nrth hidden neuron
reduce the dimensionality by EKF method
END IF
END IF

The parameters of the growing and pruning algorithm ϵ_{min}, ϵ_{max}, $S(X)$, γ and κ critically depends on the functional relationship to be approximated and also on the specified minimum approximation error (e_{min}).

4 Experiments and Discussions

In our simulations, we have used the live image quality assessment database [14]. Here, 29 JPEG images are used to generate a database of 204 JPEG images with different compression rates. including the original images, we have 233

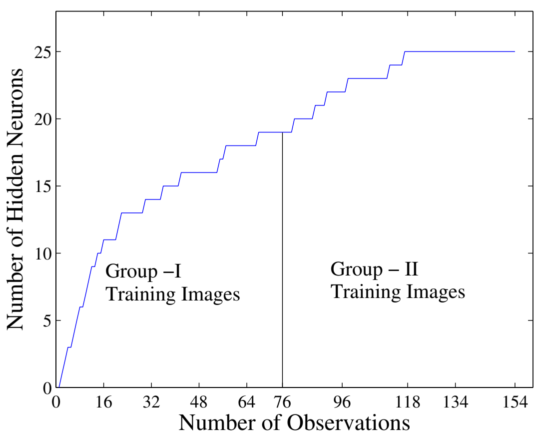


Fig. 2. Hidden neuron development

JPEG images for image quality estimation. First, the study was conducted in two sessions (first session 116 images with 20 subjects and the next session 117 images with 13 subjects). Each observer was shown the images randomly and asked to mark the perception quality on a continuous linear scale that was divided into five equal regions marked with adjectives bad, poor, fair, good and excellent. The scale was then linearly transformed to $1-10$ range. The resulting MOS was used to develop the GAP-RBF model to predict the image quality.

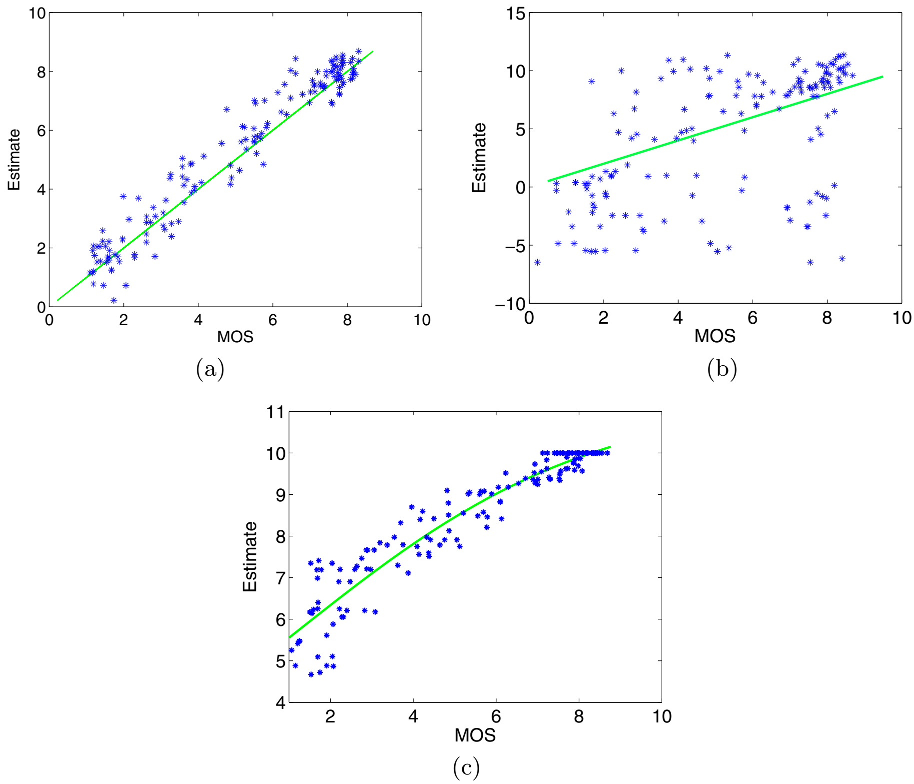
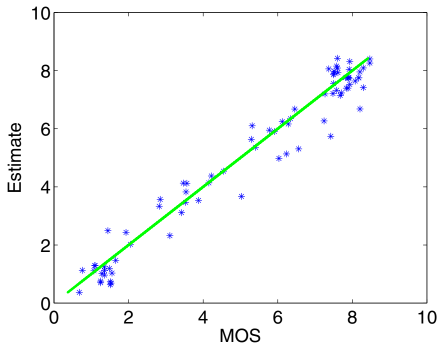
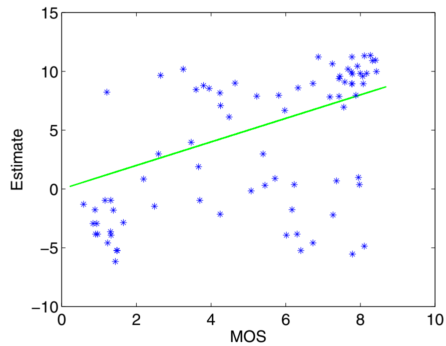


Fig. 3. Image quality prediction by (a) proposed GAP-RBF method (b) Wang-Bovik method and (c) SSIM score for 154 training images

To develop the GAP-RBF image quality model, we have selected two disjoint sets of images for training and testing. The training set images and its compressed versions are not used in testing set. Out of 29 source images, 20 images were used for training and the remaining 9 source images were used for testing. Totally 154 images were used for training (20 original and its 134 compressed versions) and 79 images for testing (9 original and its 70 compressed versions). First, we presented 154 training images sequentially to the GAP-RBF algorithm to develop the model. In our simulations, we set the following network parameters: $\epsilon_{min} = 0.001$, $\epsilon_{max} = 0.05$, $s(x) = 1$, $\gamma = 0.999$ and $\kappa = 0.1$. The expected minimum accuracy selected for our modeling is 0.0001. The GAP-RBF network initialized with zero hidden neuron, builds the network based on the 'growing and pruning' strategy mentioned earlier. The neuron history (Fig. 2) shows that 25 neurons are required to approximate the functional relationship. From Fig. 2, we see that the neuron growth saturates at 25 after the 117th training image sample. The developed GAP-RBF network model is tested with the 79 test images. The correlation between MOS and GAP-RBF based image quality metric





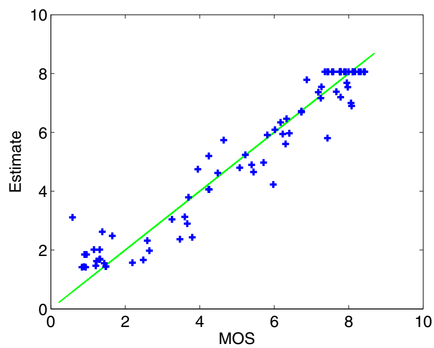


Fig. 4. Image quality prediction by (a) proposed GAP-RBF method (b) Wang-Bovik method and (c) SSIM score for 79 test images

for training and test images are shown in Figs. 3 (a) and 4 (a). Similar study is carried out using wang's NR quality metric (see Figs. 3 (b) and 4 (b)) [15] and full-reference SSIM index (Figs. 3 (c) and 4 (c)) (the SSIM index results are shown after fitting non-linear logistic function) [13]. The results clearly show that the proposed GAP-RBF model predicts the image quality better than the others. This can also be deduced from the quantitative performance analysis. The root mean square error (RMSE) deviation from MOS for image quality metric using different methods are given in table 1. From the table, it can be inferred that the proposed GAP-RBF model predicts image quality better than the other models.

Table 1. RMSE between MOS and Prediction

Metric	Testing	Training
GAP-RBF	0.46	0.61
Wang's	4.32	5.12
SSIM	0.67	0.62

5 Conclusions

In this paper, we have presented a system for predicting image quality using GAP-RBF network, considering various human visual characteristics. The functional relationship between the extracted HVS features and MOS is modeled by GAP-RBF network. Since sequential learning algorithm is used, GAP-RBF network does not require retraining when presented with a new data set. This helps us improve the model over time, receiving new sets of subjective results with minimal computational and memory requirements. The performance of the proposed metric is found to be better than other previously reported NR/FR image quality metrics.

Acknowledgment

The authors would like to thank Prof. Bovik and his lab members for providing the JPEG image quality assessment database to test our metric.

References

1. Wang, Z., Sheikh, H.R., Bovik, A.C.: No-reference perceptual quality assessment of JPEG compressed images. In: Proc. ICIP'02. Volume 1. (2002) 477–480
2. S. A. Karunasekera and N.G. Kingsbury: A distortion measure for blocking artifacts in images based on human visual sensitivity. IEEE Transactions on Image Processing **4** (1995) 713–724
3. Wu, H.R., Yuen, M.: A generalized block-edge impairment metric for video coding. IEEE Signal Processing Letters **70** (1998) 247–278
4. Suthaharan, S.: A perceptually significant block-edge impairment metric for digital video coding. In: Proc. ICASSP'2003. Volume 3., Hong Kong (2003) 681–684
5. Gao, W., Mermer, C., Kim, Y.: A de-blocking algorithm and a blockiness metric for highly compressed images. IEEE Transactions on Circuits and Systems for Video Technology **12** (2002) 1150–1159
6. Video Quality Experts Group (VQEG): (website: http://www.vqeg.org)
7. JPEG official site: (http://www.jpeg.org/)
8. Meesters, L., Martens, J.B.: A single-ended blockiness measure for JPEG-coded images. Signal Processing **82** (2002) 369–387
9. Babu, V., Perkis, A.: An HVS-based no-reference perceptual quality assessment of JPEG coded images using neural networks. In: Proc. Int. Conf. on Image Processing (ICIP'05). Volume 1., Genova, Italy (2005) 433–436
10. Karayiannis, N.B., Mi, G.W.: Growing radial basis neural networks: Merging supervised and unsupervised learning with network growth techniques. IEEE Trans. on Neural Networks **8** (1997) 1492–1506
11. Yingwei, L., Sundararajan, N., Saratchandran, P.: A sequental learning scheme for function approximation using minimal radial basis function (RBF) neural networks. Neural Computation **9** (1997) 461–478
12. G.-B. Huang, P.S., Sundararajan, N.: An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks. IEEE Trans. on Systems, Man and Cybernautics: Part-B Cybernautics **34** (2004) 2284–2292

13. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Processing **3** (2004) 600–612
14. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: (Live image quality assessment database) http://live.ece.utexas.edu/research/quality.
15. Z. Wang and A. C. Bovik and L. Lu : Why is image quality assessment so difficult? In: Proc. ICASSP'02. Volume 4. (2002) 3313–3316