https://doi.org/10.2352/issn.2169-2629.2021.29.77

The Development of Three Image Quality Evaluation Metrics Based on a Comprehensive Dataset

Dalin Tian, Muhammad Usman Khan, Ming Ronnier Luo; State Key Laboratory of Modern Optical Instrumentation, Zhejiang University; Hangzhou, China

Abstract

In this study, a large scale experiment was carried out to assess the image quality of 2266 images using categorical judgement method by 20 observers. These images were rendered in color contrast, chroma, colorfulness, lightness, and vividness directions. The results were used to derive three No-Reference (NR) Image Quality Estimation Models (IQEMs). The first model was based on color science, (different scales in CIELAB). The second model was a Neural Network model while the third model was a statistics model based on color appearance attributes. Their performances were evaluated using two databases, those developed at Zhejiang University and those available from the public databases in terms of correlation coefficients between the objective and predicted image quality scores.

Introduction

The widely spread use of mobile phones forced the manufacturers to increase the quality of their product in every aspect to stay in the competition. Mobile phone camera is one key component which is expected to produce high quality images. Typical methods include by shifting the colors of natural objects (skin, sky, grass/trees etc.) to their respective preferred color centers based on psychophysical studies, or by changing color saturation, contrast or other attributes of the perceptions. In such applications, it is necessary to evaluate the image quality, or preference. In such cases, a model of image quality is highly desired to estimate the quality of an image.

Image quality estimation can be performed either objectively (quality assessment by human observers) or subjectively (quality estimation by using computer-based algorithms). Although former is reliable to evaluate image quality, it is time-consuming and cannot be used where real time quality scores are required. So a robust image quality model is highly desired.

Subjective assessment methods can be divided into three categories namely: 1) full-reference [1-6] (to use a reference image to calculate differences against a test image to estimate image quality), 2) reduced reference (to use a reference image but only few or reduced set of features from the reference images are used for comparison) and 3) no-reference [7-17] (to perform the desired task of image quality estimation without using a reference image).

Some no-reference IQEMs focus on detecting predefined distortions present in the images like jpeg distortions detection [7] and blurring artifacts detection [8], the intensity of the detected distortions contributes inversely to the image quality, i.e. the greater intensity of the detected distortions represents lower quality of an image.

Other types of no-reference IQ estimation algorithms are based on image statistics by assuming that natural images exhibit statistical regularities and when such images are distorted, their statistics change accordingly. Moorthy *et al.* [9] used statistical models of wavelet coefficients to train a support vector regression (SVR) model to predict the severity of each distortion type present in images. Saad *et al.* [10, 11] trained probabilistic models using contrast and statistical features such as Kurtosis and anisotropy in the DCT domain.

Naturalness image quality evaluator (NIQE) by Mittal *et al*. [12] is based on constructing a collection of "Quality-Aware" features which are derived from natural scene statistics (NSS) of the pristine images. Then these features are fitted to a multivariate Gaussian (MVG) model. Image quality of a distorted image is expressed as a distance between its MVG model and MVG model of the pristine images. Integrated local NIQE proposed by Zhang *et al.* [13] is basically an improved NIQE [12] model which includes three additional quality-aware features based on statistics of gradient domain features, Log-Gabor filters and color channels in addition to the NSS features of NIQE [12]. Additionally, IL-NIQE model applied on local image patches. The pooled local quality scores were used to calculate a single quality score for an image.

The algorithm of GM-LOG-BIQA [14] by Xue *et al.* considers image local contrast features in terms of a gradient magnitude (GM) map and the Laplacian Of Gaussian (LOG) response to first employ an adaptive joint normalization procedure and then to develop their joint statistics by calculating their marginal and the conditional distributions. The calculated statistical features were used to train a support vector machine-based regression model to predict the image quality. The IQ estimation metrics discussed above mainly focused on detecting spatial deformations/abnormalities present in an image. This implies that they are not suitable for estimating the quality of images in color domain.

Choi *et al*. [15] carried out a psychophysical study using two large displays (a reference and a test display) to predict quality of color domain modified images. They proposed an image quality model based on contrast, colorfulness and naturalness attributes. It was aimed at predicting the change in image quality between the reference and test displays. Gong *et al*. [16] conducted another psychophysical experiment on smartphone and tablet displays under different lighting environments. The results were used to develop a comprehensive model which can predict image quality for different viewing conditions. The model considers naturalness and clearness attributes for determining the quality of natural scenes and colorfulness and clearness attributes for determining the quality of non-natural scenes. Although the above models [15, 16] performed well on the limited number of evaluated images, they both require a reference image to determine the image quality.

In the author's previous study [17], a no-reference image quality metric for tone-mapped images was proposed which relies on image appearance attributes of brightness and naturalness (naturalness was calculated using the attributes of colorfulness, contrast and shadow details) to predict the quality of tone-mapped images. Although it performed reasonably on the evaluated tone-mapped images, the model was incomprehensive to predict the image quality, e.g. lack of testing images and of psychophysical database testing.

The goals of the present study were to conduct a comprehensive experiment to duplicate the earlier study. The results can be analyzed to understand the reliability of visual data. Furthermore,

the data were used to develop and test three types of image quality models.

Datasets

To evaluate the performance of the presented IQEMs, we used the most comprehensive dataset. This dataset consists of two datasets developed at Zhejiang University and three publicly available datasets. We call the datasets as ZJU1 and ZJU2 datasets which were developed at Zhejiang University. The ZJU1 dataset consists of 19 original images rendered with 8 color modifications (activity, chroma, clarity, depth, heaviness, lightness, sharpness and vividness) applied at 5 levels, while the ZJU2 dataset consist of 28 original images rendered with 6 color modifications (chroma, clarity, depth, lightness, sharpness and vividness) applied at 5 levels. Together they made up for the major portion of the images, i.e. 760 images for the ZJU1 and 840 images for the ZJU2 datasets.

Three datasets (CSIQ [18], TID2013 [19] and CID:IQ [20]) were collected for IQEMs evaluation. Note only those images rendered using color attributes were used from these datasets. For the CSIQ [18] dataset, images modified by reduction of the color contrast were selected. It consists of 30 original images with contrast decrements for 4 levels. The subsets of mean shift, contrast change and color saturation rendered images were selected from the TID2013 [19] dataset. All the 24 original images were used and one laboratory image was discarded. For the TID2013 dataset, renderings of mean shift, contrast change and color saturation were made at five levels on each original image, together they make 360 images but 40 color saturation rendered images were removed due to color saturation inversion. For the CID:IQ [20] dataset, Minimum DE and SGCK Gamut mapped images were selected for use in evaluation. The CID:IQ consists of 23 original images and gamut mapping was applied at 5 levels on each image. This made up for 230 test images. In total, 2266 images form the above-mentioned datasets that were used for IQEM evaluation, including 760, 840, 116, 320 and 230 images for the ZJU1, ZJU2, CSIQ, TID2013 and CID:IQ datasets, respectively.

Psychophysical Experiment

The images from the mentioned datasets in previous section were originally evaluated differently, using different devices, using different evaluation criteria and under different surround conditions. So, a psychophysical experiment under the same viewing conditions and by using a similar display was conducted.

A NEC PW272 display with maximum luminance level set to 287 cd/m^2 was used. Images were displayed in dark surround with a neutral gray background. The distance between the observer and the monitor screen was fixed at about 80 cm (this roughly correspond to 22^o of viewing field on the display).

Fig. 1 shows experimental situation. A 6-point categorical judgement method was applied for evaluation, where the goal was to force the observers to choose good or bad ratings for displayed images instead of selecting neutral scores for some difficult to decide images. Observers were asked to select image quality rating based on the displayed image's perceptual quality using a slider whose value ranges from -3 to +3. Here, -3 corresponds to lowest image quality whereas +3 corresponds to highest image quality.

In total, 29 observers (20 Males and 9 Females, ages ranged from 22 years to 36 years) participated in the experiment where all the observers performed the Ishihara test for color deficiency prior to the experiment. Due to large number of images, the experiment was divided into 4 sessions, where each session had roughly equal

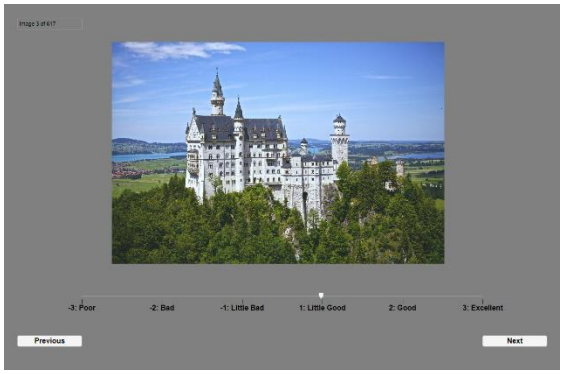


Figure 1. *Experimental window*

number of images, i.e. 620. In other words, not all the observers participated in all sessions. The instructions were given to all the observers before starting the experiment about judging the image quality in terms of ratings.

During the evaluation, 10% of the images were repeated in each session to help to understand the observer's performance. In total, 49,840 ((2266 + 10% repeat) *20 observers) image quality ratings were obtained from average 20 observers. The raw data ranged from -3 to +3 was first shifted and then normalized to the range 0 to 1. The data were then used in training and testing image quality.

Experimental Results

Intra- and inter-observer variability were first calculated using STRESS [21] measure. For intra-observer variability, repeated images data were used to calculate STRESS values between the observer's image quality ratings. The worst observer had STRESS=45, the best observer had STRESS=14 and an average value of 23. For inter-observer variability, each observer's image quality ratings were compared with the average observer ratings. The worst observer had STRESS=33, the best observer had STRESS=16 and an average value of 22. The large STRESS values for the worst observers explain the challenging nature of the task of image quality evaluation even by human observers.

The subjective image quality scores obtained from the psychophysical experiment were also compared with the scores of the original datasets. Fig. 2 shows the present psychophysical data being plotted against the original data of the five datasets respectively. Table 1 shows their correlations and STRESS values. The small scatterings can be attributed to the use of small display devices and similar group of observers at ZJU using mobile phone display in the ZJU1 and ZJU2 datasets. Note that the present evaluation method is different from the pair-comparison method used in the TID2013 database. It can be seen that there is a consistently higher score for the ZJU2 and CID:IQ datasets scores, whereas in the CSIQ datasets, the images get lower scores than their original experiment scores.

Color Science based Image Quality Model

The first approach was to develop the image quality model based on using the color appearance attributes related to image quality. However, some image appearance attributes were proved to not have consistent performance over databases. For instance, in the Gong's [16] dataset, chroma was an important attribute in modelling the image quality, but unimportant for the other database. Four more image attributes, lightness, vividness, depth and clarity were also

Table 1: Correlations and the STRESS values between the New and Old psychophysical data

Database	R	STRESS
ZJU1 (760 images)	0.65	17
ZJU2 (840 images)	0.87	12
CSIQ (116 images)	0.92	14
TID2013 (320 images)	0.79	14
CID:IQ (230 images)	0.91	17

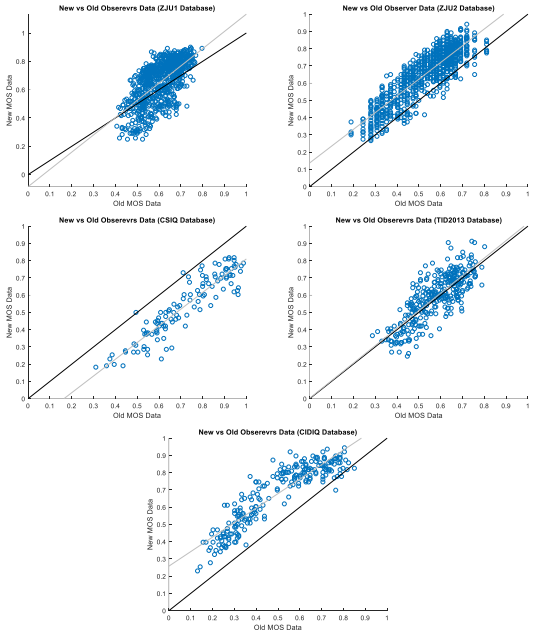


Figure 2. *New experiment data plotted against Old/Original data of the databases*

analyzed with overall image quality in each individual database. The attributes were calculated by following equations. It was found that none of above attributes had a correlation coefficient over 0.4.

$$vividness = \sqrt{L^2 + c^2}$$

$$depth = \sqrt{(L-100)^2 + c^2}$$

$$clarity = \sqrt{(L-50)^2 + c^2}$$

This rather poor performance could be explained in the following reasons.

- Overall image quality would not have obvious correlations with each single attribute since the rendering methods used in these databases are complicated and interfere with each other.
- The rendering attributes are very much image dependent and when combining together, the trend will be lost and good models would be difficult to include these attributes.
- The absolute attribute to evaluate the image quality is difficult due to the maximum chroma in different color regions are different.
- From the analysis above, the gamut boundary must be considered for every pixel. Thus, a set of new relative attributes are proposed here corresponding to the gamut ratio, including lightness ratio, chroma ratio, depth ratio, vividness ratio, and clarity ratio.

For certain attributes, the computational procedure is to increase this attribute of the target pixel until it reaches the gamut boundary. Then the end point on the gamut could be regard as a reference point. The ratio of target pixel and the end point's

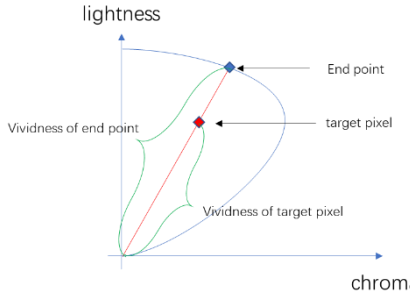


Figure 3. *The calculation procedure of vividness ratio. The blue curve is the gamut boundary in the hue page of the target pixel*

attribute will be considered as the relative attribute value. An apparent advantage of using the attribute ratio is that it always ranges from 0 to 1, which can align the diverged attribute values for different images. Also, for different pixel the correlated gamut boundary is depending on its hue, therefore the attribute of the pixels in different hue can be fairly compared. Fig. 3 explains how to acquire the end point and the vividness ratio as an example.

In addition to the above mentioned five relative attributes, four more image attributes were chosen to build the image quality model including global contrast, local contrast, naturalness, and sharpness. The unmentioned attributes were calculated in following equations.

$$naturalness = \frac{1}{1+e^{(\Delta E' - \alpha)}}$$

$$\Delta E' = k_1 * \sqrt{(l-l_0)^2 + k_2 * (a-a_0)^2 + k_3 * (b-b_0)^2 + k_4 * (a-a_0)(b-b_0)}$$

$$global\ contrast = mean(50\%\ highest\ lightness) - mean(50\%\ lowest\ lightness)$$

$$local\ contrast = the\ average\ lightness\ difference of\ each\ pixel\ and\ its\ 5 * 5\ neighbour\ pixels$$

$$sharpness = the\ average\ CIELAB\ colour\ difference of\ each\ pixel\ and\ its\ 5 * 5\ neighbour\ pixels$$

All nine attributes were considered to fit the model in the first stage. However, it is not necessary to use all nine attribute pairs because of the attributes having analogous calculation method, for instance lightness ratio and global contrast, chroma ratio and clarity ratio, local contrast and sharpness, etc.

To reduce the effects of cross terms and simplify the model, those attributes which had least contribution to image quality was removed. To rank the contribution of all attributes, each attribute was removed in turn and the correlation coefficient and STRESS of the model with rest of the terms were calculated. A small decrease in correlation (or a small increase in STRESS) represents not a significant decrease in model accuracy if this attribute is removed, indicating the attribute's lower contribution in predicting image quality. Table 2 shows the correlation coefficients (R) and the STRESS values. The value in each grid is the prediction accuracy when remove this attribute.

To balance the model accuracy and model complication, 6 attributes were finally preserved including naturalness, global contrast, local contrast, clarity ratio, vividness ratio and sharpness. Following equation shows the final model.

$$IQ = -0.33 * vividness\ ratio + 0.43 * clarity\ ratio + \cdots 0.23 * global\ contrast - 50.27 * local\ contrast + \cdots 0.18 * naturalness + 0.009 * sharpness + 0.5$$

Table 2: The correlation coefficients and STRESS of the rest model when removing an attribute

Number of attributes	naturalness	global contrast	local contrast	clarity ratio	vividness ratio	sharpness	depth ratio	lightness ratio	chroma ratio	R/STRESS
9	0.56/18.6	0.62/17.5	0.54/18.8	0.63/17.3	0.65/17.1	0.57/18.3	0.64/17.2	0.65/16.9	0.66/16.8	0.66/16.8
8	0.54/18.8	0.62/17.5	0.54/18.8	0.53/19.0	0.65/17.1	0.57/18.3	0.62/17.4	0.65/16.9		0.66/16.8
7	0.54/18.8	0.59/18	0.52/19.0	0.52/19.1	0.57/18.4	0.57/18.4	0.61/17.8			0.65/16.9
6	0.48/19.5	0.51/19.2	0.48/19.6	0.51/19.3	0.54/18.8	0.54/18.8				0.61/17.8
5	0.45/19.9	0.44/20.1	0.47/19.7	0.45/19.9	0.51/19.3					0.54/18.8
4	0.44/20.1	0.43/20.2	0.43/20.2	0.44/20.1						0.51/19.3
3	0.34/21.1	0.31/21.3	0.39/20.6							0.45/20.1
2	0.27/21.6	0.29/21.4								0.39/20.6

Neural Network based Image Quality Model

The second approach is based on Convolutional Neural Network (CNN) which we trained to predict the image quality. Transfer learning was used from GoogLeNet [22]. We replaced the original input R, G, B by 8 color related elements for each pixel, including L*, a*, b*, chroma, hue, vividness, depth and clarity, to concentrate on the color performance of an image instead of edge detection. All images in database were resized to 224×224 pixels, therefore the input of the network was a 224×224×8 matrix. And the output of the network was the observers' preference range from 0 to 1, as mentioned before. Since the target of the network changed from object classification to image quality prediction, the responses also changed from classes to scores, thus the output layer should be modified from a classification layer to a regression layer. The inside structure was all preserved and more details could be found in the description of GoogLeNet.

Image Joint Statistics based Image Quality Model

The third IQ model is based on the work of Xue *et. al.* [14]. Fig. 4a shows the graphical representation of the original GM- LOG –BIQA [14] model. In their work, the authors utilized two image local contrast features, namely Gradient Magnitude (GM) and Laplacian of Gaussian (LOG), to calculate their joint statistics. First, the two feature maps were normalized jointly, this makes the local image contrast scales of the two feature maps consistent across the entire image. GM and LOG features describe local image structures and interaction between them can help predict the perceptual image quality, to achieve this the authors calculated joint statistics of these feature maps in terms of marginal and conditional probabilities. Since this IQEM predict the quality of images modified in spatial domain (noise, blur, compression etc.) with higher accuracy, it's performance on images modified in color domain (contrast, mean shift, chroma, lightness etc.) is not very impressive. This motivated us to evaluate the usefulness of the color appearance attributes of the images to calculate their joint statistical features and test their performance on the images modified in color domain.

Fig. 4b shows the graphical representation of the modified GM-LOG-BIQA [14] model where Vividness-Depth (V-D) and Brightness-Colorfulness (Q-M) attribute pairs are used. To calculate the joint statistical features, the input RGB image is first transformed in to XYZ color space by using the display conversion model. Image color appearance attributes of brightness (Q) and colorfulness (M) were obtained from CAM16-UCS [23] whereas the vividness (V) and depth (D) were calculated from the following equations by using the attributes of colorfulness (M) and brightness (Q)

$$\mathrm{V,D} = \mathrm{k}_1\sqrt{(\mathrm{Q}-\mathrm{Q}_0)^2 + \mathrm{k}_2\mathrm{M}^2} \quad (1)$$

For vividness $k_1 = 0.15, k_2 = 30$ & $Q_0 = 0$ whereas for depth $k_1 = 0.4, k_2 = 1$ & $Q_0 = 300$. After obtaining the image color appearance attributes, joint adaptive normalization (JAN) is

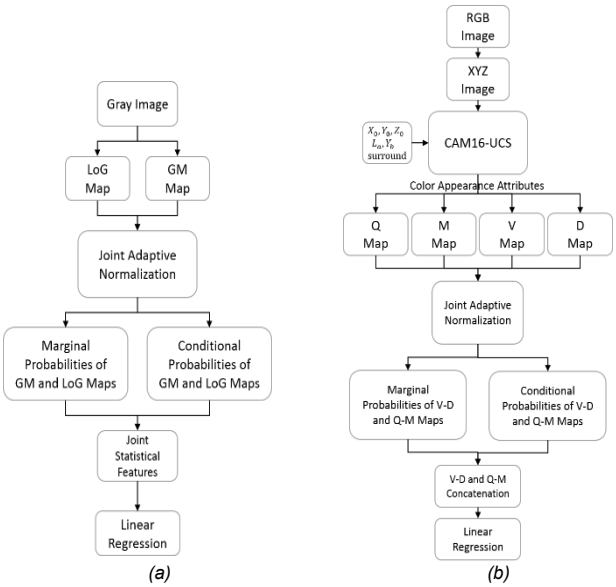


Figure 4. *(a) Design flow of the GM-LOG-BIQA [6] algorithm, b) Design flow of the proposed Image Statistical (IS) model*

applied on these attributes using the following equations

$$F_I(i,j) = \sqrt{V^2 + D^2 + Q^2 + M^2} \quad (2)$$

Then a locally adaptive normalization factor was computed at each pixel location (i,j)

$$N_I(i,j) = \sqrt{\sum\sum_{(l,k)\in\Omega_{i,j}} \omega(l,k)F_I^2(l,k)} \quad (3)$$

Where $\omega(l,k)$ is a Gaussian kernel scaled to unit sum i.e. $\sum_{l,k}\ \omega(l,k) = 1$ and $\Omega_{i,j}$ is a local window operator centered at each pixel location(i,j). Normalized features can now be calculated as

$$\bar{V} = V_I/(N_I + \varepsilon) \quad (4)$$
$$\bar{D} = D_I/(N_I + \varepsilon) \quad (5)$$
$$\bar{Q} = Q_I/(N_I + \varepsilon) \quad (6)$$
$$\bar{M} = M_I/(N_I + \varepsilon) \quad (7)$$

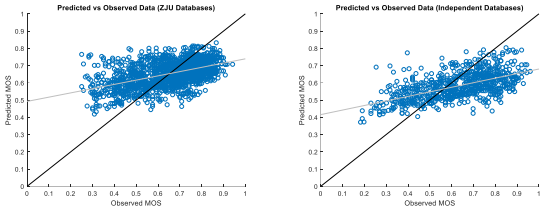
Where ε is a small constant added to avoid numerical instability. After JAN, the remaining steps to calculate marginal and conditional probabilities for V-D and Q-M attribute pairs followed the same methodology as in original GM-LOG-BIQA [14] model. We quantized the visual attributes data into 10 bins (same number of bins as used in original model [14]) for calculating probabilities. This result in two marginal probability vectors and two conditional probability vectors, each of length 10, and combining them together returns a feature vector of length 40 for each of the two attribute pairs V-D and Q-M. We finally combined the feature vectors of two visual attribute pairs (V-D and Q-M) to get a final feature vector consisting of 80 entries for a given input image. These features were then utilized in training a linear regression model to predict the quality of an input image.

IQEM's Performance Evaluation

In order to evaluate the performance of presented IQEMs, we used the datasets mentioned in the Datasets section. We trained the regression model using the calculated features from the ZJU dataset (1600 images) and tested the performance on the independent dataset (666 images). For CS and IS model's features, linear regression was used for training while SGDM optimization algorithm was utilized for training on the NN model features.

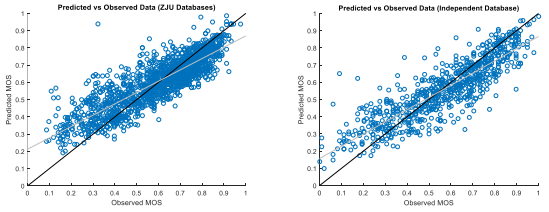
Table 3: Correlation R, STRESS, RMSE and MAE values for the training and testing datasets for the three IQEMs

Model	Data	R	STRESS	RMSE	MAE
CS	Training (1600 images)	0.50	19	0.24	0.18
	Testing (666 images)	**0.58**	**21**	**0.21**	**0.16**
NN	Training (1600 images)	0.85	16	0.11	0.08
	Testing (666 images)	**0.87**	**18**	**0.10**	**0.07**
IS	Training (1600 images)	0.79	14	0.09	0.07
	Testing (666 images)	**0.62**	**21**	**0.16**	**0.13**



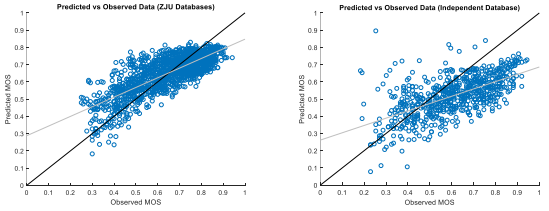
(a) (b)

Figure 5. *The prediction accuracy of color science (CS) model containing 6 terms for (a) training and (b) testing dataset*



(a) (b)

Figure 6. *Prediction accuracy of the Neural Network (NN) model for (a) training and (b) testing dataset*



(a) (b)

Figure 7. *Prediction accuracy of the Image Statistics (IS) model for (a) training and (b) testing dataset*

Table 3 reported the performance scores in terms of correlation, STRESS, RMSE and MAE values for the three IQEMs. It can be seen that NN model performed the best among the three IQEMs with respective R and STRESS scores of 0.85 and 16 for the ZJU training dataset and 0.87 and 18 for the independent testing database. Their respective RMSE and MAE values are 0.11 and 0.08 for training data while 0.10 and 0.07 for the testing data.

IS model performed second best with respective R and STRESS scores of 0.79 and 14 for the ZJU training dataset and 0.62 and 21 for the independent testing dataset. Their RMSE and MAE values are 0.09 and 0.07 for training data while 0.16 and 0.13 for the testing data, respectively. Performance of CS model was found to be the worst with respective R and STRESS scores of 0.50 and 19 for the ZJU training dataset and 0.58 and 21 for the independent testing dataset. Here, RMSE and MAE values are 0.24 and 0.18 for training data while 0.21 and 0.16 for the testing data, respectively. Figs. 5-7 show the predicted vs observed IQ scores for the three IQEMs on (a) ZJU training dataset and (b) independent testing dataset.

Conclusion

Three IQEMs based on color science (CS), neural network (NN) and image statistics (IS) were developed to predict the quality of color domain modified images. Also, a large scale experiment was carried out to collect image quality data based on a large number of previously assessed images under same viewing conditions. The visual results were used to test these models. The results showed that NN model performed the best, followed by IS and CS models.

References

[1] Z. Wang and A.C. Bovik, "Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures," IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, 2009.

[2] CIE, "Colorimetry," Technical Report, 15, 2004.

[3] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color-image reproduction," Jour. of the Society for Inf. Display, vol. 5, no. 1, pp. 61-63, 1997.

[4] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Processing, vol. 13, no. 4, pp. 600–612, 2004.

[5] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," IEEE Transactions on Image Processing, vol. 22, no. 2, pp. 657-667, 2013.

[6] H. Z. Nafchi, A. Shahkolaei and R. F. Moghammadam, "FSITM: A Feature Similarity Index for Tone-Mapped Images," IEEE Signal Processing Letters, vol. 22, no. 8, pp. 1026-1029, 2015.

[7] Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in IEEE Int. Conf. on Image Process., pp. 477-480, 2002.

[8] J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," in Signal Process. Image Commun. 19, pp. 147–161, 2004.

[9] A. K. Moorthy and A. C. Bovik. "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.

[10] M. A. Saad, A. C. Bovik and C. Charrier, "A DCT statistics-based blind image quality index," IEEE Signal Processing Letters, vol. 17, no. 6, pp. 583-586, 2010.

[11] M. A. Saad, A. C. Bovik, and C. Charrier, "Model-Based Blind Image Quality Assessment Using Natural DCT Statistics," IEEE Trans. On Image Process., vol. 21, pp. 3339–3352, 2011.

[12] 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.

[13] L. Zhang, L. Zhang and A. C. Bovik, “A feature-enriched completely blind local image quality analyzer,” IEEE Transactions on Image Processing. vol. 24, no. 8, pp. 2579-2591, 2015.

[14] W. Xue, X. Mou, L. Zhang, A. C. Bovik and X. Feng, “Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features,” IEEE Transactions on Image Processing, vol. 23, no. 11, pp. 4850-4862, 2014.

[15] S. Y. Choi, M. R. Luo, M. R. Pointer and P. A. Rhodes, “Predicting perceived colorfulness, contrast, naturalness and quality for color images reproduced on a large display,” 16th Color Imaging Conference, IS&T and SID, pp. 158-164, 2008.

[16] R. Gong, H. Xu, M. R. Luo and H. Li, “Comprehensive model for predicting perceptual image quality of smart mobile devices,” Appl. Opt. vol. 54, pp. 85-95, 2015.

[17] M. U. Khan, I. Mehmood, M. R. Luo, M. F. Mughal, “No-Reference Image Quality Metric for Tone-Mapped Images,’ in 27th Color Imaging Conference, IS&T and SID, pp. 252-255, 2019.

[18] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," J. Electron. Imaging 19, 011006, Jan 07, 2010.

[19] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, C.-C. Jay Kuo, Image database TID2013: Peculiarities, results and perspectives, Signal Processing: Image Communication, vol. 30, pp. 57-77, 2015.

[20] X. Liu, M. Pedersen, and J.Y. Hardeberg, “CID:IQ - a new image quality database,” in Image and Signal Processing, A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds., vol. 8509 of Lecture Notes in Computer Science, pp. 193–202. Springer, Cherbourg, France, Jul. 2014.

[21] P. A. García, R. Huertas, M. Melgosa, and G. Cui, “Measurement of the relationship between perceived and computed color differences,” J. Opt. Soc. Am. A vol. 24, pp. 1823–1829, 2007.

[22] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-9, 2015.

[23] C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. H. Brill and M. Pointer, “Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS,” Color Research & Application, vol. 42, no. 6, pp. 703-718, 2017.