IMPROVING THE PERFORMANCE OF NO-REFERENCE IMAGE QUALITY ASSESSMENT ALGORITHM FOR CONTRAST-DISTORTED IMAGES USING NATURAL SCENE STATISTICS

Yusra Al-Najjar[1] and Chen Soong Der[2]

ABSTRACT

This study was conducted to explore the role of two color features in improving the performance of the existing No-Reference Image Quality Assessment Algorithms for Contrast-Distorted Images (NR-IQA-CDI). The used color features were Colorfulness and Naturalness of color expressed in CIELab and CIELuv color spaces. Test images used were the public benchmark databases that contain contrast-distorted images - TID2013, CID2013 and CSIQ. Experiments for the exploration were conducted in two stages: the preliminary stage and the comprehensive stage. The results of preliminary study showed that the features of colorfulness and naturalness of color can improve the prediction of human opinion score which relies mainly on the feature of brightness-only contrast. The results inspired to more comprehensive study where the Natural Scene Statistics (NSS) of these two features were estimated by modelling the probability distribution function (pdf) of 16,873 test images from a public database called SUN2012. The results based on k-fold cross validation with k ranging from 2 to 10 showed that the performance of NR-IQA-CDI can be improved by adding the NSS of these features.

1. INTRODUCTION

Image quality assessment (IQA) is an important study area in image processing and computer vision [1], since it requires assessing the performance of various image processing algorithms. Human Visual System (HVS) is the ultimate receiver and interpreter of image content; therefore, subjective assessment is considered the most reliable assessment method. However, subjective assessment is time-consuming, expensive and requires a lot of effort. To overcome this limitation, many Image Quality Assessment algorithms (IQAs) have been proposed over the past decade. The aim of IQA is to predict image quality in a manner that is consistent with the results of subjective assessment [2]-[3].

The IQAs available are classified into three categories according to the level of access to the reference image - the distortion - free or perfect-quality image - which are: Full-Reference IQAs (FR-IQAs) in which there is full access to the reference image, Reduced-Reference IQAs (RR-IQAs) which have access to only some of the information about the reference image [4]-[5] and the No-Reference IQAs (NR-IQAs) that require no information about any reference image [6]. In many applications, where there is no information about the reference image, the NR-IQA is highly desired [1]. One of such applications is the assessment of contrast-distorted image (CDI). Contrast distortion happens during image acquisition [1], where acquisition devices are not perfect or lighting is poor. This might cause loss of contrast and small details. In such case, this acquired image will be the original image, but it cannot be used as a reference image with perfect quality.

In general, contrast change – distortion – is an important aspect in the field of image evaluation [1]. Nevertheless, despite all the work that has been proposed, there is still a lack for an algorithm that evaluates a contrast-distorted image that has no reference. This study aims to explore some elements that affect the evaluation of contrast distorted image by using the NSS of the color features.

1.   Y. Al-Najjar is with the Faculty of Information and Communication Technology, Universiti Tenaga Nasional, Malaysia. Email: yusra.najjar.2011@gmail.com
2.   Chen S. Der is with the Faculty of Information and Communication Technology, Universiti Tenaga Nasional, Malaysia. Email: sc2339@gmail.com

Recently image quality assessment algorithms have been proposed dealing with different types of distortion, such as compression, blur or noise, but little with contrast distortion. No one can claim that there is an IQA algorithm that is perfectly consistent with human opinion. Some of them were close to subjective assessment, but for one database only. So, studies are still going on to find a better IQA algorithm. To help with this, in our research, we focused on studying elements of the algorithm – basic bricks - instead of building the assessment algorithm and the main target database was TID2013 database, which recent IQA algorithms failed to evaluate [1]. Depending on what others have reached in this field, we decided to start working on improving the performance of these NR-IQA-CDI algorithms. And to do that, we needed to specify new features as will be explained in this research.

Following sections include a comprehensive review of the work carried out on NR-IQA algorithms, research methodology that included brief explanation of colorfulness and naturalness of color, performance metrics used in the experiments, databases used, statistical tests and procedure of the experiments, followed by an evaluation of results, and finally a conclusion.

2. LITERATURE REVIEW

When NR-IQA algorithms started, they were not based on specific types of distortion, which means that they were general image quality assessment algorithms detecting distortion in general; but, what if the distortion was specific: compression, blur, noise or contrast alone. Some of the algorithms were specialized to assess images with one distortion type such as blur [7], whereas others were general algorithms. Our main concern in this research is contrast distortion. Currently, it is noticed that general NR metrics usually follow two main thrusts; being natural scene statistics (NSS)-based or learning-based. NSS-based metrics extract properties from the image using statistical properties in the natural image, while learning-based metrics are involved in learning and testing based on neural networks or support vector regression (SVR). These two types are not quite different, where NSS-based algorithms assume a certain statistical system in the spatial domain of natural scene [8].

In [9], Gu et al. proposed a new no-reference (NR)/ blind sharpness metric in the autoregressive (AR) parameter space. This metric was established by analyzing AR model parameters. They calculated the energy and contrast differences in the locally estimated AR coefficients in a pointwise way, then they quantified image sharpness with percentile pooling to predict the overall score. They validated the effect of the technique on subsets of blurring artefacts from four large-scale image databases (LIVE, TID2008, CSIQ and TID2013). Despite the claim that the metric was good, but it was not tested using contrast-distorted images.

In [10], the authors proposed a metric based on Curvelet No-Reference Transform (CNR), which outperformed other full reference metrics, such as SSIM and PSNR, in predicting the level of distortion noise, JPEG compression or blur in natural images. But, this metric did not take into consideration contrast change (distortion); neither global nor local. Later in 2010, [11], Lua et al. proposed a no-reference metric that used contourlet transform based on an improved model of NSS - CNSS – established with contourlets. They claimed that this algorithm was superior to the conventional NSS model and could be applied to any distortion. This algorithm was general and was applied only to LIVE database. It was not tested on global or local contrast-distorted images. Li et al. in [12] used phase congruency, entropy and gradient as image features to assess image quality depending on the general regression neural network (GRNN). Li et al. used LIVE database and divided it into five datasets. Five-fold cross-validation was used. This metric was a general metric, concentrated on one database and did not consider contrast change, whether global or local.

Distortions change the statistical properties found in natural images. Moorthy et al. in [13] proposed a blind IQA framework and integrated algorithm based on natural scene statistics called DIIVINE index. This metric did not compute specific distortion features, but it extracted statistical features. This metric dealt with compression, noise, blur and fading. The metric was trained over LIVE database and evaluated using only TID2008. DIIVINE metric did not deal with contrast change and was limited to one global enhanced database for testing. In 2011, researchers started to work on contrast-distorted images. In [14], the authors claimed that the contrast quality is determined by two metrics; the histogram flatness (HF) and the Histogram Spread (HS). They claimed that low-contrast images have a low HS value, whereas high-contrast images have a high HS value. This means that HS can

differentiate between low- and high-contrast images. The images used were natural and medical images. The HS metric can specify whether the image requires more enhancement or not, but not specifying its quality. This metric was not tested on databases; besides, this metric relies on partial access to the reference image, which makes it a reduced reference metric and not a no-reference metric.

There have been NR metrics that were designed for general purpose usage. The authors of [15] and [16] built statistical models of mean subtracted contrast normalized (MSCN) coefficients and spatial relationship between neighbouring pixels. The model was trained on features obtained from both natural and distorted images and on human judgments of the quality of these images. Therefore, BRISQUE metric in [15] was limited to the types of distortion it has been tuned to. It worked fine for noise, blur and compression distortion, but its performance degraded significantly for contrast-distorted images, whereas for [16] there was a limited number of databases, and by comparison, the NIQE Index was not tied to any specific distortion type.

In 2013 [17], Gu et al. proposed a reduced reference metric for contrast-change images called RIQMC metric depending on information residual between the input and the distorted images as well as the first four order statistics of the distorted image histogram. Gu et al. used CID2013, TID2013 and CSIQ public databases. This RIQMC metric was devised based on phase congruency and information statistics of histogram, acquiring superior performance beyond existing models and managing to enhance original natural images. Despite that the metric achieved an impressive performance, the major drawback was that there was partial access to the reference image and that access was not based on natural image statistics (NSS) model. It inevitably needs a single number (the entropy) from original natural image. Entropy is considered an important feature, while it cannot represent the local information of the image. Xue et al. used statistics associated with gradient magnitude and Laplacian features to measure image quality [18]. Xue et al. proposed a blind image quality metric BIQA that predicts image quality by analyzing the image statistics in some transformed domains, such as discrete cosine transform domain and wavelet domain. Used databases were LIVE, TID2008 and CSIQ. Despite that the results for BIQA were good, it dealt with distortion in general such as blur and compression, but not with contrast-distortion.

In 2015 [1], Fang et al. proposed a no-reference image quality metric for contrast-distorted images depending on natural scene statistics (NSS). They employed many images to build an NSS model based on moment features. The authors used the three public databases CID2013, TID2013 and CSIQ. The results of their experiments were good for one of the databases, but not for TID2013 database, which called for more research.

Li et al. in [7] used discrete orthogonal moments to evaluate blur effect. They proposed a blind blur assessment metric concentrating on blur distortion only. They used four public databases, which are LIVE, CSIQ, TID2008 and TID2013, in their experiments. The authors claimed that blur affects the magnitudes of moments of an image based on discrete Tchebichef moments. The gradient of a blurred image is first computed to account for the shape, which is more effective for blur representation. Then, the gradient image is divided into equal-size blocks and the Tchebichef moments are calculated to characterize the image shape. The energy of a block was computed as the sum of squared non-DC moment values. Finally, the proposed image blur score is defined as the variance-normalized moment energy, which is computed with the guidance of a visual saliency model to adapt to the characteristics of human visual system. This metric attempts to model indicators of quality for the distortion in question – blur - and hence was unsuitable for use for contrast change.

In [19], Liu et al. proposed a no-reference image quality assessment method depending on mutual information of wavelet domain named (MIQA-II), by computing neighbouring pixels using LIVE [20] database. This method showed good results in identifying distortion. But, this method was based only on LIVE database without taking contrast change into consideration. In [21], Gu et al. reported a new large dedicated contrast-changed image database (CCID2014) which includes 655 images and associated subjective ratings recorded from 22 inexperienced observers and proposed a reduced reference image quality metric for contrast change (RIQMC) using phase congruency and statistics information of the image histogram. Validation of the proposed model was conducted on contrast related to CCID2014, TID2008, CSIQ and TID2013 databases. Despite that the metric showed good results, it required information from the original image – which was entropy – in addition to that it was

not tested on locally enhanced images. Depending on this metric, Wu et al. in [22] proposed a no-reference metric for contrast-distorted image assessment. They extracted five statistical features from the distortion image and two features were extracted from the phase congruence (PC) map of distortion image. These features and human mean opinion scores (MOS) of training images were jointly utilized to train a model of Support Vector Regression (SVR). Despite that the results were close to RIQMC results, these results depended on one database only which is CCIID2014. In [23], Gu et al. presented a no-reference image quality assessment metric for contrast-distorted images. In the metric, they searched for local details. They first removed predicted regions in an image, claiming that the unpredicted ones are of much information. Then, they computed entropy of unpredicted areas of maximum information by visual saliency. Gu et al. used CID2013, CCID2014, CSIQ, TID2008 and TID2013 databases. According to the results achieved, there was still weakness in predicting quality, especially for TID2013, where the results was 0.64.

Lately, Shokrollahi et al. proposed a contrast-changed image quality (CCIQ) metric including a local index, named edge-based contrast criterion (ECC) and three global measures [24]. They did not consider the metric as a full-reference metric, since the original image is not regarded to have the ideal quality. They claimed that it follows a new paradigm in image quality assessment. Experimental results on the three benchmark databases CID2013, TID2013 and TID2008 demonstrate that the proposed metric outperforms the state-of-the-art methods. This metric showed good evaluation for CID2013 and TID2013 when compared to full-reference metric (PSNR), reduced reference metric (RIQMC, [17]) and no-reference metric (NR-CDIQA, [1]). Regardless of the results, this metric is still referring to the original – reference - image.

In [25], Gu et al. developed a blind/no-reference (NR) model for assessing the perceptual quality of screen content pictures with big data learning. In this model, they extracted four types of features descriptive of picture complexity, screen content statistics, global brightness quality and sharpness of details. The efficacy of the new model was compared with existing blind picture quality assessment algorithms applied on screen content image databases. The proposed model gave a promising performance.

In [26], Gu et al. investigated the problem of image quality assessment (IQA) and enhancement using machine learning. In their work, they developed a new NR-IQA model by extracting 17 features from the given image by analyzing contrast, sharpness, brightness, among other features. They validated the efficiency of their metric using nine datasets. Another contribution of the authors was image enhancement to be based on quality optimization. The authors conducted histogram modifications to modify image brightness and contrast to a proper level. They claimed that their framework can enhance image contrast and lightness.

3. RESEARCH METHODOLOGY

Experiments for exploring colorfulness and naturalness features were conducted in two stages: the preliminary stage, that explores the features of contrast and color using the raw data; this stage was fast and easy. Depending on the results from the preliminary stage, the comprehensive stage - using the NSS of the features - was conducted. The comprehensive stage required more time and work. The study implementation of this study started with:

    1- Getting the distribution of local contrast, colorfulness and naturalness for SUN2012 database using *dfittool( )* in MATLAB software.
    2- Getting the probability distribution function (pdf) of these features.
    3- The implementation was conducted over three public databases CID2013 [27], [17], TID2013 [28] and CSIQ [29].
    4- The regression method used was Support Vector Regression (SVR) with cross-validation for k = 2 to 10.
    5- The performance metrics and tests used are mentioned later in sub-sections 3.3 and 3.5.

3.1 Colorfulness Feature

Colorfulness is a low-level feature of color; it is a visual sensation by which the perceived colorfulness of any part of an object appears to be less or more chromatic [30]. Images were converted CIELab

205

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 3, No. 3, December 2017.

color space which contains all perceivable colors. CIELab was used, because it has an infinite range of colors that exceeds those in RGB color model to simulate human vision. In addition, it is device-independent. See Figure 1.

Figure 1. To the right is the image in RGB and to the left is the image in CIELab color space.

3.2 Naturalness Feature

Naturalness is a high-level feature of color; it is composed of hue, saturation, chroma and luminance, which makes it a good complement to colorfulness. For naturalness computation, images were converted into CIELuv color space that uses previously mentioned color components.

3.3 Performance Metrics Used in the Experiments

For measuring the correlation with MOS, Pearson Linear Correlation Coefficient (PLCC) and Root Mean Squared Error (RMSE) were used to predict accuracy. For monotonicity, Spearman Rank Order Correlation Coefficient (SROCC) was used.

3.4 Databases

Databases used in the experiments were: first the CID2013 database which was established for contrast-distorted images. It contains 15 references and 400 contrast-enhanced images. Twenty-two people have shared in providing quality score for MOS. Second, the TID2013 database that contains 25 reference images with different distortion types including contrast distortion. The last was the CSIQ database with 30 reference images and different types of distortion. CSIQ database contains 116 contrast-distorted images. Subjective tests on the three databases followed the recommendations of ITU-R BT 500-12 [31]. A sample of images used in the research is displayed in Figure 2.

Figure 2. Sample of original (to the left) and contrast-distorted images. Top row from CSIQ database, middle row from CID2013 database and bottom row from TID2013 database.

3.5 Statistical Test

Paired *t*-test was conducted for each of the four performance metrics based on the metric's value obtained before and after modification in either feature or regression. The output of t-test was the p-value; the probability that the differences between two groups of data in a pair were not statistically significant; the higher the probability, the more likely the differences are not statistically significant. In

this study, the *p*-value less than 0.05 was interpreted as the differences being statistically significant. The results were reported in the form of percentage of difference for each of the databases and average differences over all databases.

3.6 Procedure of the Experiments

Conducted experiments in the preliminary stage included:

1. Computing local contrast, colorfulness and naturalness for each image in the three public databases.
2. Preparing the five moment features for each database (mean, standard deviation, entropy, skewness and kurtosis) for color features.
3. Using Support Vector Regression (SVR) for predicting the output data to be compared with target data of MOS.
4. Computing PLCC, SROCC and RMSE correlations performance metrics.
5. Computing t-test for each performance metric.

3.6.1 Preliminary Stage

Contrast vs. (Contrast + Colorfulness) and Contrast vs. (Contrast + Naturalness)

This experiment measures the correlation between MOS and local contrast feature before and after adding each of color features - colorfulness and naturalness - to the contrast.

1. Image data were partitioned using cross-validation *cvpartition*( ) function for k-fold from 2 to 10 to minimize bias.
2. In each round – from the k sets – each set acts as a test set only once, whereas the remaining are training sets and same for all sets.
3. PLCC, SROCC and RMSE performance metrics were used to find the correlation between MOS and local contrast.
4. Each k group prediction was averaged.
5. Percentage of difference before and after adding the (raw or NSS) of the feature was computed.
6. The statistical parametric t-test was performed to test the null hypothesis.

3.6.2 Comprehensive Stage

(Five moment features) vs. (five moment features + pdf of Colorfulness feature) and (five moment features) vs. (five features + pdf of Naturalness feature)

To compute colorfulness of color, the image was converted from RGB color space into CIELab color space, then Chroma of the color was computed according to the equation:

$$C^* = (a^{*2} + b^{*2})^{0.5}$$
<div align="right">Equation (1)</div>

where C* is the computed Chroma. Colorfulness of color was computed according to the following equation:

$$Ck = mean(C^*) + std(C^*)$$
<div align="right">Equation (2)</div>

The steps to compute the naturalness of color were to convert the image from RGB color space into CIELuv color space. Images were first converted from RGB color space into XYZ color space, then the parameters for Luv were computed according to the equations:

$$L^* = \begin{cases} 116(\frac{Y}{Y_n})^{\frac{1}{3}} - 16 \ if \ \frac{Y}{Y_n} > \ 0.008856 \\ 903.3\left(\frac{Y}{Y_n}\right) \quad if \ \frac{Y}{Y_n} \le 0.008856 \end{cases}$$
<div align="right">Equation (3)</div>

$$u^* \ = \ 13 \ (L^*)(\acute{u} - \acute{u}_n)$$
<div align="right">Equation (4)</div>

$$v^* = 13 \, (L^*)(\acute{v} - \acute{v}_n) \qquad \qquad \text{Equation (5)}$$

$$\text{where } \acute{u} = \frac{4X}{X+15Y+3Z} \, , and \, \acute{v} = \frac{9Y}{X+15Y+3Z} \qquad \text{Equation (6)}$$

L^* scales from 0 to 100.

Then the computed variables of luminance, hue and saturation were used in computing naturalness of color for each image as follows:

$$C^* = (u^{*2} + v^{*2})^{0.5} \qquad \qquad \text{Equation (7)}$$

$$h_{uv} = \arctan(\frac{v^*}{u^*}) \qquad \qquad \text{Equation (8)}$$

$$S_{uv} = \frac{C^*}{L^*} \qquad \qquad \text{Equation (9)}$$

Finally, the naturalness of color was computed according to the equation:

$$N_{ij} = \exp(-0.5(\frac{S_{ij} - \mu_j}{\sigma_j})^2) \qquad \qquad \text{Equation (10)}$$

This experiment measures the correlation between MOS and the probability of the five features (mean, standard deviation, entropy, skewness and kurtosis) before and after adding the probability of each feature of colorfulness and naturalness. Steps followed were the same as in the preliminary experiment.

4. EVALUATION OF RESULTS

The results of the preliminary stage were encouraging for moving to the comprehensive stage. So, the results displayed here concern the comprehensive stage only.

Table 1 and Table 2 summarize the results. Table 1, row 1 shows the percentage of difference in each of the performance metrics for adding the colorfulness feature. Table 2, row 1 shows the p-values of the paired t-tests. As seen in Table 1, there was a good improvement in the results of the experiment when using TID2013 database – this database was a target for improvement. PLCC and SROCC increased by 3.04% and 4.92% and RMSE decreased slightly by 0.94%. As seen in Table 2, all the three p-values for TID2013 database were less than 0.05, indicating that the differences in all three performance metrics were statistically significant.

As for CID2013, there were marginal decrements in PLCC and SROCC by 0.36% and 0.59%, respectively, with a marginal increment in RMSE by 1.08%. As seen in Table 2, all the three p-values for CID2013 were less than 0.05, indicating that the differences in all the three performance metrics were statistically significant. Nevertheless, it is worth noting that the magnitudes of change in the three metrics were very marginal.

For CSIQ database, there were good increments in PLCC and SROCC by 13.69% and 13.42%, respectively, while there was a good decrement in RMSE by 11.44%. The statistical test results in Table 2 indicated that the differences in the three performance metrics were statistically significant.

For the average results over the three databases, there were good increments in PLCC and SROCC by 5.46% and 5.92%, respectively, with a decrement in RMSE by 3.77%. The results of statistical tests showed that there were statistically significant improvements in PLCC, SROCC and RMSE, since the p-values were less than 0.05. Overall analysis indicated that adding colorfulness could improve the performance of NR-IQA-CDI.

This was for adding colorfulness feature. As for adding naturalness feature, Table 1, row 2 shows the percentage of difference in each of the performance metrics, whereas Table 2, row 2 shows the p-values of the paired t-tests. As seen in Table 1, there was an improvement in the results of experiment using TID2013 database. PLCC increased by 3.53%, while SROCC slightly increased by 0.17%. RMSE decreased by 1.42%. As seen in Table 2, the p-values for PLCC and RMSE were less than 0.05, indicating that the differences in these two performance metrics were statistically significant.

However, there was no statistically significant difference in SROCC.

As for CID2013 database, there were slight increments in PLCC and SROCC by 0.43% and 0.73%, respectively, with a marginal decrement in RMSE by 0.8%. As seen in Table 2, all the three p-values for CID2013 were less than 0.05, indicating that the differences in all three performance metrics were statistically significant.

For CSIQ database, there were increments in PLCC and SROCC by 17.05% and 18.67%, respectively and there was a decrement in RMSE by 12.88%. The statistical test results indicated that the differences in all three performance metrics were statistically significant.

For the average results over the three databases, there were increments in PLCC and SROCC by 7.00% and 6.53%, respectively and a decrement in RMSE by 5.03%. The results of statistical tests showed that there were statistically significant differences in PLCC, SROCC and RMSE, because the p-values were less than 0.05. Overall analysis indicated that adding naturalness could improve the performance of NR-IQA-CDI.

Coming to adding both colorfulness and naturalness together, results were displayed as follows:

Table 1, row 3, shows the percentage of difference in each of the performance metrics and Table 2, row 3, shows the p-values of the paired t-tests. As seen in Table 1, there was an improvement in the results of experiment using TID2013 database, which was a target for improvement. PLCC and SROCC increased by 5.19% and 5.28%, respectively. RMSE decreased by 1.79%. As seen in Table 2, for all three performance metrics - PLCC, SROCC and RMSE - p-values for TID2013 were less than 0.05, indicating that the differences in all three performance metrics were statistically significant.

As for CID2013 database, there was a slight decrement in PLCC by 0.07%, whereas SROCC slightly increased by 0.19%, with a marginal decrement in RMSE by 0.67%. As seen in Table 2, SROCC and RMSE p-values for CID2013 were less than 0.05, indicating that the differences in these performance metrics were statistically significant.

For CSIQ database, there were good increments in PLCC and SROCC by 19.18% and 20.50%, respectively. RMSE showed a good decrement by 15.47%. The p-values for statistical test results were less than 0.05, indicating that the differences in the three performance metrics were statistically significant.

Averaging the results over the three databases, there were increments in PLCC and SROCC by 8.10% and 8.65%, respectively, with a decrement in RMSE by 5.53%. The results of statistical tests showed that there was statistically significant improvement in PLCC, SROCC and RMSE, because all the p-values were less than 0.05. Overall, the results indicated that adding both colorfulness and naturalness of color could help improve the performance of NR-IQA-CDI.

Comparing adding each color feature alone to adding both together gave the results listed in Table3 and Table 4. Table 3 shows the percentage of difference in each of the performance metrics and Table 4 shows the p-values of the paired t-tests for adding colorfulness or naturalness versus adding both.

Table 3 shows that PLCC increased by 2.42% and 1.04% when adding both features as compared to adding only colorfulness and adding only naturalness, respectively. The p-values of statistical tests were less than 0.05, indicating that the differences in the performance metrics for adding both features were statistically significant.

SROCC increased by 2.48% and 2.03% when adding both features as compared to adding only colorfulness and adding only naturalness, respectively. The p-values of statistical test results were less than 0.05, indicating that the differences in the performance metrics for adding both features were statistically significant.

RMSE decreased by 1.95% and 0.59% when adding both features as compared to adding only colorfulness and adding only naturalness, respectively. The p-value of statistical test for both features versus colorfulness only was less than 0.05, indicating that the differences in the performance metrics were statistically significant. There was no statistically significant difference for adding both features versus adding only naturalness, but this has only a small effect.

Table 1. Percentage of difference in the performance after adding natural scene statistics of both colorfulness and naturalness to the five features each one separated and both together.

Feature/Database / Performance Metric	TID2013			CID2013			CSIQ			AVERAGE		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Colorful-ness	3.04%	4.92%	-0.94%	-0.36%	-0.59%	1.08%	13.69%	13.42%	-11.44%	5.46%	5.92%	-3.77%
Natural-ness	3.53%	0.17%	-1.42%	0.43%	0.73%	-0.80%	17.05%	18.67%	-12.88%	7.00%	6.53%	-5.03%
Both Features (Colorfulness & Naturalness)	5.19%	5.28%	-1.79%	-0.07%	0.19%	0.67%	19.18%	20.50%	-15.47%	8.10%	8.65%	-5.53%

Table 2. P-values for the performance for adding natural scene statistics of both colorfulness and naturalness to the five features each one separated and both together.

Feature /Database / Performance Metric	TID2013			CID2013			CSIQ			AVERAGE		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Colorful-ness	0.0019	0.0002	0.0030	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003
Natural-ness	0.0000	0.3060	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000
Both Features (Colorfulness & Naturalness)	0.0001	0.0001	0.0001	0.1480	0.0482	0.0016	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 3. Percentage of difference in the performance after adding both features vs. colorfulness and both features vs. naturalness over the three public databases.

Features	PLCC	SROCC	RMSE
Both vs. Colorfulness	2.42%	2.48%	-1.95%
Both vs. Naturalness	1.04%	2.03%	-0.59%

Table 4. *P*-values for paired t-test on the difference in the performance of NR-IQA-CDI after adding both features vs. adding only one of them.

	PLCC	SROCC	RMSE
Both vs. Colorfulness	1.6454×10^{-06}	7.5066×10^{-05}	1.1542×10^{-09}
Both vs. Naturalness	2.3174×10^{-03}	6.5035×10^{-04}	2.1236×10^{-01}

5. CONCLUSION

This study tested the hypotheses that: "adding colorfulness of color could improve the performance of NR-IQA-CDI in predicting MOS" and "adding naturalness of color could improve the performance of NR-IQA-CDI in predicting MOS".

Results denoted that there was a significant positive difference – improvement in the performance of prediction – for adding only colorfulness of color feature, for adding only naturalness of color feature and even for adding both of color features together in evaluating the no-reference image quality. Adding both color features versus adding either one of them showed – in general – that adding both features gives better results.

The variety of images in each database and the number of images played a part in the results. But, in spite of this, the results came promising for the proposed features in predicting the quality of images that have no reference.

REFERENCES

[1]   Y. Fang, K. Ma, Z. Wang and W. Lin, "No-Reference Quality Assessment of Contrast-Distorted Images Based on Natural Scene Statistics," IEEE Signal Processing Letters, vol. 22, no. 7, pp. 838-842, 2015.

[2]   W. Lin and C. -C. J. Kuo, "Perceptual Visual Quality Metrics: A survey," Visual Communication and Image Representation, vol. 22, no. 4, pp. 297-312, 2011.

[3]   A. C. B. Zhou Wang, Modern Image Quality Assessment (Synthesis Lectures on Image, Video, & Multimedia Processing), San Mateo, CA, USA: Morgan & Claypool Publishers, 2006.

[4]   Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Processing, CA, 2004.

[5]   H. R. Sheikh and A. C. Bovik, "Image Information and Visual Quality," Transactions on Image Processing, Montreal, Que., Canada, 2006.

[6]   R. Hassen, Z. Wang and M. Salama, "No-Reference Image Sharpness Assessment Based on Local Phase Coherence Measurement," The IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 2010.

[7]   L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami and A. C. Kot, "No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments," in IEEE Transactions on Cybernetics, Cybern, 2016.

[8]   D. L. Ruderman and W. Bialek, "Statistics of Natural Images: Scaling in the Woods," Physical Review Letters, vol. 73, no. 6, pp. 551-558, 1995.

[9]   K. Gu, G. Zhai, W. Lin, X. Yang and W. Zhang, "No-Reference Image Sharpness Assessment in Autoregressive Parameter Space," Shanghai Municipal Commission of Economy and Informatization, China, Shanghai, 2015.

[10]   J. Shen, Q. Li and G. Erlebacher, "Curvelet-Based No-Reference Objective Image Quality Assessment," Proceedings of the 27th Picture Coding Symposium (PCS'09), NJ, USA, 2009.

[11]   W. Lua, K. Zing, D. Tao, Y. Yuan and X. Gao, "No-Reference Image QualityAassessment in Contourlet Domain," Neurocomputing, vol. 73, no. 4-6, pp. 784-794, January 2010.

[12]   C. Li, A. C. Bovik and X. Wu, "Blind Image Quality Assessment Using a General Regression Neural Network," IEEE Transactions on Neural Networks, 2011.

[13]   A. K. Moorthy and Alan Conrad Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," IEEE Transactions on Image Processing, 2011.

[14]   A. K. Tripathi, S. Mukhopadhyay and A. K. Dhara, "Performance Metrics for Image Contrast," International Conference on Image Information Processing (ICIIP), Shimla, India, Nov. 2011.

[15]   A. Mittal, A. K. Moorthy and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," IEEE Transactions on Image Processing, August, 2012.

[16]   A. Mittal, R. Soundarajan and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," IEEE Signal Processing Letters, pp. 209-212, 2013.

[17]   K. Gu, G. Zhai, X. Yang, W. Zhang and M. Liu, "Subjective and Objective Quality Assessment for Images with Contrast Change," Image Processing (ICIP), Melbourne, VIC, Australia, Sept. 2013.

[18]   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, 2014.

[19]   B. Liu, L.-X. Liu, H.-P. Dong and Y.-G. Lin, "Blind Image Quality Assessment Based on Mutual Information," Electronics, Communications and Networks, vol. 382, pp. 127-136, 29 June 2016.

[20]   H. Sheikh, Z. Wang, L. Cormack and A. Bovik, "LIVE Image Quality Assessment Database Release 2," [Online]. Available: http://live.ece.utexas.edu/research/quality. [Accessed 2014].

[21]   K. Gu, G. Zhai, W. Lin and M. Liu, "The Analysis of Image Contrast: From Quality Assessment to Automatic Enhancement," Transactions on Cybernetics (T-)CYP, vol. 46, no. 1, pp. 284 - 297, Jan 2016.

[22]   J. Wu, Z. Xia, Y. Ren and H. Li, "No-Reference Quality Assessment for Contrast-Distorted Image," International Conference on Image Processing Theory Tools and Applications (IPTA), Oulu, Finland, Dec. 2016.

[23]   K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang and C. W. Chen, "No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization," IEEE Transactions on CYBERNETICS, 2016.

[24]   A. Shokrollahi, A. Mahmoudi-Aznaveh and B. M.-N. Maybodi, "Image Quality Assessment for Contrast Enhancement Evaluation," International Journal of Electronics and Communications (AEÜ), pp. 61-66, 2017.

[25]   K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin and A. C. Bovik, "No-Reference Quality Assessment of Screen Content Pictures," IEEE Transactions on Image Processing, China, Singapore, 2017.

[26]   K. Gu, D. Tao, J.-F. Qiao and W. Lin, "Learning a No-Reference Quality Assessment Model of Enhanced Images with Big Data," IEEE Transactions on Neural Networks and Learning Systems, Singapore, 2017.

[27]   T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen and J. Häkkinen, "CID2013: A Database for Evaluating No-Reference Image Quality Assessment Algorithms," IEEE Transactions on Image Processing, vol. 24, no. 1, pp. 390-402, 2015.

[28]   N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola and B. Vozel, "Color Image Database TID2013: Perculiarities and Preliminary Results," The 4th European Workshop on Visual Information, 2013.

[29]   E. C. Larson and D. M. Chandler, "Categorical Image Quality (CSIQ) Database," 2010. [Online], Available: http://vision.okstate.edu/csiq.

[30]   M. Faichild, Color Appearance Models, John Wiley & Sons, p. 87, 2013.

[31]   I. R. Assembly, "Methodology for the Subjective Assessment of the Quality of Television Pictures ITU - Radiocom Sector," International Telecommunication Union, 2013.

"Improving the Performance of No-Reference Image Quality Assessment Algorithm for Contrast-Distorted Images Using Natural Scene Statistics", Yusra Al-Najjar and Chen Soong Der.

ملخص البحث:

أجريـت هـذه الدراسـة بهـدف استكشـاف دور اثنتـين مـن خصـائص اللـون فـي تحسـين أداء خوارزميـات تقيـيم جـودة الصـور عديمـة المرجـع القائمـة بالنسـبة للصـور مشـوهة التبـاين. هاتـان الخاصـيتان همـا: حيويـة اللـون وطبيعتـه. أمـا الصـور التـي تـم اختبارهـا فكانـت مـن قواعـد بيانـات مرجعيـة تشـتمل علـى صـور مشـوهة التبـاين. وتـم إجـراء التجـارب فـي مـرحلتين همـا: المرحلـة الأوليـة، والمرحلـة الشـاملة. وقـد بينـت نتـائج المرحلـة الأوليـة مـن الدراسـة أن خاصـيتي حيويـة اللـون وطبيعتـه يمكنهمـا تحسـين التوقـع الخـاص بـالرأي البشـري الـذي يرتكـز علـى خاصـية التبـاين المـرتبط بشـدة الإضاءة فقط.

وقـد شـجعت نتـائج المرحلـة الأوليـة علـى القيـام بدراسـة شـاملة مـن أجـل تقـدير إحصـائيات المشـهد الطبيعـي للخاصـيتين المـذكورتين، وذلـك عبـر نمذجـة دالـة توزيـع الاحتمـال لِـ 16873 صـورة اختبـار مـن قاعـدة البيانـات SUN2012 . وأوضـحت النتـائج المبنيـة علـى التأييـد التقـاطعي المضـاعف، حيـث يتـراوح معامـل المضـاعفة مـن 2 إلـى 10، أن أداء خوارزميـات تقيـيم جـودة الصـور عديمـة المرجـع يمكـن تحسـينه بإضـافة إحصـائيات المشهد الطبيعي لخاصيتي حيوية اللون وطبيعته إلى الصور المختبرة.