ABSTRACT

Most No-Reference Image Quality Assessment (NR-IQA) metrics are designed for the quality assessment of images distorted by compression, noise and blurring. Few NR-IQA metrics exist for Contrast-Distorted Images (CDI). Reduced-reference Image Quality Metric for Contrast-changed images (RIQMC) and NR-IQA for Contrast-Distorted Images (NR-IQACDI) are the state-of-the-art IQA algorithms for CDI. Room for improvement exists, especially for the assessment results using the image database called TID2013. The current NR-IQACDI uses features in spatial domain. This paper proposes the use of the same statistical features but in Curvelet domain, which is powerful in capturing the multiscale and multidirectional information of an image. Experiments are conducted to assess the effect of using statistical features in Curvelet domain. The experimental results are based on K-fold cross validation with K range from (2 to 10). The statistical tests indicate that the performance using selected statistical features in the Curvelet domain are better than that of the NR-IQACDI. The use of other statistical features and selection methods should be further investigated to increase the prediction performance.

Keywords: image quality assessment (IQA), no reference image quality assessment (NR-IQA), contrast distortion images (CDI), multiscale geometric analysis (MGA) transforms, curvelet domain.

1. INTRODUCTION

The images may be distorted during various processes, e.g. acquisition, processing, compression, storage, transmission, reproduction and sharing. In order to measure the image quality, approaches for IQA or Video Quality Assessment (VQA) such as subjective and objective methods have been proposed [1].

Subjective quality assessment is impractical for real-time applications because it is time-consuming and expensive. Therefore, Objective IQA algorithms are preferable because they can analyze the images and predict the quality without human role. Depending on the availability of an “ideal quality” original image, objective IQAs can be further classified into Full Reference (FR), Reduced Reference (RR) and No Reference (NR) [2]. The general taxonomy of IQA/VQA is shown in Figure-1 [3].

Loss of contrast and visible details in an original image may be attributed to the limitation of the acquisition device and poor lighting condition [4]. Therefore, the acquired image is the original source and hence the perfect-quality reference is unavailable in this case. As such, No Reference Image Quality Assessment Algorithm (NRIQAs) can be used to optimize the settings of the Contrast Enhancement (CE) algorithms by providing the quality of the output contrast changed images.

Most of the existing NR-IQA metrics focus on the quality evaluation of distorted images due to compression, noise, and blurring. It is worth to mention here that the related work performed in the area of NR-IQA for CDI is quite limited. Figure-2 illustrates the reference image with contrast distorted types [5].

However, most existing state-of-the-art IQA algorithms are dedicated for CDI based on statistical features such as mean, variance, skewness, kurtosis, and entropy of pixel intensities [6,7,5]. In [6], the reduced-reference image quality metric for contrast-changed images (RIQMC) was based on the entropies and order statistics of the image histograms. In [7], a no-reference (NR) IQA method for contrast enhancement was proposed based on the principle of natural scene statistics (NSS). In [5], contrast quality was determined by the histogram flatness and spread. The performance of the state-of-the-art NR-IQACDI still requires improvement. Therefore, in this work, instead of relying on the statistical features of the spatial domain, we first decompose an image to its subbands using the Curvelet transform with different scale and direction. We then extract statistical features for each coefficient in each subband. Statistical features such as mean, variance, skewness, kurtosis and entropy are obtained and selected from both spatial and Curvelet domains. The feature selection method is applied to select the most significant features.
In previous literature, multiscale Geometrical Analysis (MGA) transforms had a significant function in IQA. Typically, MGA transforms [8] increase the number of transforms by combining both multiscale and multidirectional transform properties. MGA transforms include the following: ridgelets, curvelets, wave atoms, contourlets, shearlet, steerable pyramid, etc. Liu et al. [9], Wen et al. [10], Li et al. [11] and Fang et al. [12] built the NR-IQA metric based on NSS of the curvelet, contourlet, shearlet and steerable pyramid decomposition domains, respectively.

MGA methods have a significant role in contrast image applications [13,14,15,16,17], and contrast has a significant influence on the quality of an image in human visual perception. However, the methods missed its role in the research of IQA for CDI.

Most of the existing NR-IQA metrics are based on MGA transforms and are designed for the assessment of the quality of image distorted by compression, noise and blurring. However, no NR-IQA measures based on MGA transforms that have been specifically developed for CDI exist [18].

Curvelet transform is a special member of MGA transform. Their higher directional sensitivity, anisotropy, and lesser redundancy allow them to represent edges and other singularities along curves more efficiently than traditional transforms [19]. Given the inherited advantages of the curvelet transform, it is widely used in image processing applications such as contrast enhancement [13]. Natural distortion-free images are known to possess specific statistical properties and that distortions may change these properties. Based on this idea, [20] proposed the first NSS-based NR-IQA method in the curvelet domain. Thereafter, [9] and [21] introduced a new NR-IQA metric based on the curvelet transform and NSS. [9] proposed a good NR-IQA metric from their experimental results. Their results exhibited a set of energy features extracted in the Curvelet domain that were highly relevant to natural image quality across multiple distortion categories. Their findings are among the best examples of NSS and can accurately predict image quality.

Experiments are conducted to assess the effect of using statistical features in Curvelet domain. Results based on K-fold cross validation with K range from (2 to 10) and statistical test indicate that performance using selected statistical features in Curvelet domain could be better than that of NRIQACDI. The use of other statistical features and selection methods should be further investigated to increase quality prediction performance. In the next section (Section 2), the Curvelet transform is described. Section 3 describes the procedure of the proposed method, including feature extraction and selection, and the quality

Figure-1. IQA measurement classifications [3].

Figure-2. (a) Original ‘lenna’ image, (b) low contrast (dark) ‘lenna’ image, (c) low contrast (bright) ‘lenna’ image, (d) histogram equalized ‘lenna’ image [5].
assessment. Section 4 describes the experimental results, and Section 5 concludes the current work.

2. CURVELET TRANSFORM

The Curvelet transform, which is one of the members in Multiscale Geometrical Analysis (MGA) transform, is designed to better represent edges and other singularities along curves via the implementation of an effective parabolic scaling law: width $\cong$ (length)$^2$ on the sub-bands appeared in the frequency domain. Curvelet transform is well known for its higher directional sensitivity, higher anisotropy and lesser redundancy [19]. Figure-3 shows the edge representations by both Wavelet and Curvelet Transforms.

Candes et al. [22] proposed two fast discrete curvelet transforms (FDCT), the unequally-spaced fast fourier transform (USFFT) - based curvelet and frequency wrapping based curvelet. Both methods differ in terms of the choice of the spatial grid used to translate the curvelet at each scale and angle. Note that both digital transformations return a table of digital curvelet coefficients indexed by scale, orientation, and location parameters. FDCT is simpler, faster and less redundant than the first generation. In this paper, we apply FDCT via wrapping with six levels.

The discrete Curvelet transform of a 2-D function $f(t_1, t_2)$ is defined as follows:

$$C^D(j, \ell, k) = \sum_{0 \leq t_1, t_2 < n} f(t_1, t_2) \psi_{j, \ell, k}^D(t_1, t_2),$$  \hspace{1cm} (1)

Where $\psi, j, \ell, \text{ and } k$ are Curvelet functions, scale, orientation and position respectively. $t_1, t_2$ denote coordinates in the spatial domain: $0 \leq t_1, t_2 < n$. $C^D(j, \ell, k)$ denotes Curvelet coefficient. The two parameters involved in the digital implementation of the Curvelet transform are the number of resolution and the number of angles at the coarsest level. The parameters are bounded by two constraints, the maximum number of resolutions depending on the original image size and the number of angle at the second coarsest level, which must be at least eight and multiple of four.

3. PROPOSED METHOD

In this study, curvelet transform coefficients are used in the score prediction process. Here, we discuss the feature extraction and selection and the quality assessment. The details of each step are elaborated in the following sections. Figure-4 depicts the procedure.

![Figure-3. Edge Representations by both Wavelet and Curvelet Transforms. Left side: Take many Wavelet coefficients to accurately represent such a curve that means it take lots of memory. Right side: Curvelets can represent a smooth contour with much fewer nonzero coefficients for the same precision [22].](image)

![Figure-4. Block diagram of proposed method.](image)
3.1 Feature extraction and selection

In this study, a set of feature vector is extracted from different domains in three contrast distorted image databases, i.e. CID2013, TID2013, and CSIQ. In Curvelet domain, decomposition of the distorted image into six-scale curvelet coefficients at different scales and orientations is achieved via the fast discrete curvelet transform (wrapping). The numbers of matrix of coefficient for levels 1, 2, …, 6 are 1, 16, 32, 32, 64 and 1, respectively. Hence, the total number of matrices is 146. Five statistical features such as mean, standard deviation, entropy and entropy are calculated for each level.

In spatial domain, statistical image features such as mean, variance, skewness, kurtosis and entropy are extracted directly from each distorted image. In order to improve the prediction rate, Sequential Forward Search (SFS) [23, 24] is applied to select a minimal number of the optimal features that are sensitive to contrast distortion. The selected feature vector is entire to score prediction step as shown in the next section.

3.2 Image quality assessment

The quality of the final selected feature vector is subsequently predicted. Apart from classification problems, Support Vector Machines (SVM) can be adopted to solve regression problems as well. Here, Support Vector Regression (SVR) is used to identify the mapping function $f(x)$ between the feature set $x_i$ and the subjective quality score. Then, the mapping function is trained to predict quality scores using SVR. The LIBSVM-3.12 package [25] is used to build the regression model, which is mainly used to predict human MOS from final NSS features vector.

The whole training data set is represented as $(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)$, where vector $x_i$ is the $i^{th}$ sample of the feature and $y_i$ is the corresponding subjective quality score target for $i=1, 2, 3, ..., N$. Our next aim is to find the mapping function $\hat{y}_i = f(x_i)$ that predict $y_i$. The function can be determined by follows:

$$\hat{y} = f(x) = w^T \psi(x) + \gamma$$  \hspace{1cm} (2)

where $\psi(x)$ is a kernel function of the feature vector $x$, which is used to implicitly map the problem from lower to a higher dimensional space. Several popular kernels include the linear, sigmoid, radial basis function (RBF), and polynomial functions. In the current data training process, we use the default setting of RBF in [25], where $w$ is the weighting vector and $\gamma$ is the bias term. The job of SVR is to estimate $\psi$, $w$ and $\gamma$ in (2). In the testing stage, the test feature vector $X_j$ of the jth test image is serves as the input to the system to create the objective score $\hat{y}_j$.

4. EXPERIMENTAL PROCEDURE AND RESULTS ANALYSIS

4.1 IQA Databases

For the experiments, we select the test images from the three publicly available databases, the CSIQ [26], TID2013 [27] and CID2013 [6]. We use only the contrast distorted images in the three databases (that is, reference images are excluded). A total of 116, 250, and 400 distorted images are selected from CSIQ [26], TID2013 [27], and CID2013 [6], respectively. The distorted image sizes for CSIQ, TID2013, and CID2013 were 512 x 512 pixels, 384 x 512 pixels, and 768 x 512 pixels, respectively. The difference Mean Opinion Scores (DMOS) associated with distorted images, which is ranging from 0 to 1, is reported, where a lower DMOS signifies a higher quality. We performed experiments on a laptop with an Intel (R) Core (TM) 2 Duo CPU, 2G RAM memory with a MATLAB R2013a platform.

4.2 The IQA performance metrics

We use three performance metricsto assess the performance of IQA: the Spearman rank-order correlation coefficient (SROCC), Pearson’s (linear) correlation coefficient (PLCC), and the root mean square error (RMSE) among the predicted objective scores and the subjective mean opinion scores (MOS). A good performance in terms of its correlation with human perception is normally indicated by SROCC, LCC~1, and RMSE~0. These metric measures the prediction monotonicity, prediction accuracy, and prediction consistency.

Given that regression is essentially a learning algorithm that requires training, K-fold cross validation (CV) was used for the assessment of the performance of IQA to assess how well the IQA could be generalized to independent groups of data while minimizing bias.

While performing K-fold cross validation, three databases are randomly partitioned into 10 subsets and 10-fold leave-one-out cross-validation is used to test the enhanced proposed metric. Here, 90 % of the database is used for training set while the rest are used for testing set. Assessment is repeated for K rounds and the results are averaged. In order to reduce variability, multiple rounds of cross-validation (k = 2 to 10) are performed on different partitions. In order to avoid bias, the above cross-validation is repeated 100 times and the results are averaged as shown in Table-1 and Table-2.
Table-1. The average PLCC, SROCC and RMSE across 100 train-test rounds for three Databases using features in spatial domain.

<table>
<thead>
<tr>
<th>k</th>
<th>CSI Q</th>
<th>TID2013</th>
<th>CID2013</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLCC</td>
<td>SROCC</td>
<td>RMSE</td>
</tr>
<tr>
<td>2</td>
<td>0.6358</td>
<td>0.6165</td>
<td>0.1312</td>
</tr>
<tr>
<td>3</td>
<td>0.6317</td>
<td>0.5947</td>
<td>0.1310</td>
</tr>
<tr>
<td>4</td>
<td>0.6285</td>
<td>0.6191</td>
<td>0.1314</td>
</tr>
<tr>
<td>5</td>
<td>0.6559</td>
<td>0.6346</td>
<td>0.1297</td>
</tr>
<tr>
<td>6</td>
<td>0.6382</td>
<td>0.5972</td>
<td>0.1352</td>
</tr>
<tr>
<td>7</td>
<td>0.6625</td>
<td>0.6362</td>
<td>0.1248</td>
</tr>
<tr>
<td>8</td>
<td>0.6358</td>
<td>0.6184</td>
<td>0.1223</td>
</tr>
<tr>
<td>9</td>
<td>0.6458</td>
<td>0.5647</td>
<td>0.1262</td>
</tr>
<tr>
<td>10</td>
<td>0.5890</td>
<td>0.5426</td>
<td>0.1303</td>
</tr>
</tbody>
</table>

Table-2. The average PLCC, SROCC and RMSE across 100 train-test rounds for three Databases using features in Curvelet domain.

<table>
<thead>
<tr>
<th>k</th>
<th>CSI Q</th>
<th>TID2013</th>
<th>CID2013</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLCC</td>
<td>SROCC</td>
<td>RMSE</td>
</tr>
<tr>
<td>2</td>
<td>0.8729</td>
<td>0.8642</td>
<td>0.0884</td>
</tr>
<tr>
<td>3</td>
<td>0.8367</td>
<td>0.8306</td>
<td>0.0905</td>
</tr>
<tr>
<td>4</td>
<td>0.8955</td>
<td>0.8743</td>
<td>0.0780</td>
</tr>
<tr>
<td>5</td>
<td>0.8974</td>
<td>0.8706</td>
<td>0.0768</td>
</tr>
<tr>
<td>6</td>
<td>0.8698</td>
<td>0.8460</td>
<td>0.0733</td>
</tr>
<tr>
<td>7</td>
<td>0.8788</td>
<td>0.8510</td>
<td>0.0773</td>
</tr>
<tr>
<td>8</td>
<td>0.8508</td>
<td>0.7642</td>
<td>0.0747</td>
</tr>
<tr>
<td>9</td>
<td>0.8854</td>
<td>0.8191</td>
<td>0.0773</td>
</tr>
<tr>
<td>10</td>
<td>0.9005</td>
<td>0.8256</td>
<td>0.0777</td>
</tr>
</tbody>
</table>

Table-1 shows the average result of assessment using the feature vector obtained from spatial domain (NRIQACDI). Table-2 shows the average result of assessment using the feature vector obtained from Curvelet domain. The result in Table-2 are better than the result in Table-1. The performance improved when the Curvelet transform was used. The next section discusses and identifies whether the differences in the performances among NRIQACDI and selected statistical features in curvelet domain are significant.

4.3 Statistical performance analysis

Let \( c_i \) is performance metric values of NRIQACDI and \( cvt c_i \) is performance metric values of using Curvelet features. Then we calculate the difference between the two performance metrics for each \( k \) in each of the databases.

\[
d_i = cvt c_i - c_i
\]  

Then the average of the percentage of differences of all the \( k \) values and databases are computed. The percentage is measured by dividing the difference in performance by the absolute value of performance metric of \( c_i \)

\[
dp = \frac{1}{n} (\Sigma_{i=1}^{n} d_i / abs(c_i))
\]  

where \( n \) is the number of all \( k \) in all databases. The absolute value is used to preserve the sign of difference of performance in the percentage (increment or decrement). Table-3 shows the percentage of difference in each of the performance metrics.
Table-3. Differences results for NRIQACDI - Statistical Curvelet features and using feature selection.

<table>
<thead>
<tr>
<th></th>
<th>PLCC</th>
<th>SROCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TID2013</td>
<td>48.63%</td>
<td>41.77%</td>
<td>-25.55%</td>
</tr>
<tr>
<td>CID2013</td>
<td>9.81%</td>
<td>9.02%</td>
<td>-31.36%</td>
</tr>
<tr>
<td>CSIQ</td>
<td>37.98%</td>
<td>39.38%</td>
<td>-38.54%</td>
</tr>
<tr>
<td>Overall Databases</td>
<td>32.14%</td>
<td>30.06%</td>
<td>-31.82%</td>
</tr>
</tbody>
</table>

In order to evaluate the statistical significance of performance of each metric, hypothesis testing based on the Paired T-tests is applied on the performance metric value obtained by NRIQACDI and Curvelet features in order to produce the p-value (see Table-4). In general, p-value of < 0.05 shows that there is a significant difference within the values.

Table-4. P-Values for paired T-TEST shows superiority of results during use five features in Curvelet domain.

<table>
<thead>
<tr>
<th></th>
<th>PLCC</th>
<th>SROCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TID2013</td>
<td>6.162E-09</td>
<td>1.3751E-07</td>
<td>1.6038E-07</td>
</tr>
<tr>
<td>CID2013</td>
<td>3.569E-06</td>
<td>1.6807E-05</td>
<td>1.4733E-06</td>
</tr>
<tr>
<td>CSIQ</td>
<td>8.4711E-09</td>
<td>3.8977E-08</td>
<td>5.7208E-09</td>
</tr>
<tr>
<td>Overall Databases</td>
<td>2.0777E-12</td>
<td>6.6974E-12</td>
<td>3.4357E-09</td>
</tr>
</tbody>
</table>

4.4 Discussions

Based on the experiments results in Table-3 and Table-4, we can clarify the following:

- For the TID2013 database, which is one of our target for improvement, a good increment was observed in PLCC and SROCC, with ratings of 48.63% and 41.77%, respectively. The RMSE has a good decrement of 25.55%. All the three P-values for TID2013 in Table-4 are less than 0.05, indicating that significant differences exist in all three performance matrices.
- As for CID2013, the PLCC and SROCC increased by 9.81% and 9.02%, respectively, and the RMSE decrement by 31.36%. All the three P-values for CID2013 in Table-4 are less than 0.05, indicating that significant differences exist among the three performance matrices.
- For the CSIQ database, which is our secondary target for improvement, a good increment was observed in PLCC and SROCC, with ratings of 37.98% and 39.38%, respectively. Furthermore, a good decrement of 38.54% was observed in the RMSE. The statistical results indicate that significant differences exist among the three performance matrices.
- For the average results over the three databases, a good increment in PLCC and Spearman was observed, with ratings of 32.14% and 30.06%, respectively. Furthermore, a decrement in RMSE of 31.82% was observed. All the three P-values for overall databases in Table-4 are less than 0.05, indicating that significant differences exist among all the three performance matrices.

The results of statistical tests show significant improvement in PLCC, Spearman, and RMSE because the P-values are less than 0.05. Overall analysis indicates that using selected statistical features in Curvelet domain could have a better performance than that of the NRIQACDI.

5. CONCLUSIONS

Of course, the accuracy can be improved by the effective use of multiple features of various domains and the proper selection of a suitable feature selection method. The main objective of the current paper is to assess the effect of using statistical features in Curvelet domain. The experiment results based on K-fold cross validation (with K ranging from 2 to 10) and statistical test indicate that the performance by using selected statistical features in Curvelet domain is better than that of NRIQACDI. The improvement in prediction accuracy may be attributed to the sparse representation of the Curvelet transform. Of course, other statistical features and feature selection methods can be considered to improve the performance. The related study will be reported in the future work.

REFERENCES


