IMAGE QUALITY ASSESSMENT - A SURVEY OF RECENT APPROACHES

Noor Al Madeed, Zainab Awan and Somaya Al Madeed

College of Computer Science and Engineering, Qatar University, Doha, Qatar

ABSTRACT

Image Quality Assessment (IQA) is the process of quantifying degradation in image quality. With the increased image-based applications, IQA deserves extensive research. In this paper we have presented popular IQA methods for the three types namely, Full Reference (FR), No Reference (NR) and Reduced Reference (RR). The paper gives comparison of the approaches in terms of the database used, the performance metric and the methods used.

KEYWORDS

Full reference, image quality assessment, no reference, reduced reference

1. INTRODUCTION

Image quality assessment refers to the evaluation of the quality of a distorted image with respect to the original image. There are two main categories of IQA; namely, subjective methods and objective methods. The former is difficult and time-consuming for real-time assessment needs. Subjective methods require human intervention and are very expensive. The later refers to the algorithmic models to estimate the image quality. For the real-time assessment we require automated or objective methods. There are three dimensions of objective IQA, namely, Full Reference FR, Reduced Reference RR and No Reference NR. FR IQA methods need the whole information of the reference image, while NR IQA methods stipulate that distorted image be assessed without any information of the reference image. RR methods are a compromise between FR and NR methods that require partial information of the reference image.

The rest of the paper has been organised into following sections. Section 2 defines image quality, section 3 explains image quality assessment, section 4 outlines types of IQA, and section 5 highlights few benchmark databases, followed by section 6 that discusses IQA approaches, followed by the discussion and conclusion.

2. IMAGE QUALITY

Image quality has been defined [1] as a composite of three factors, fidelity, perception and aesthetics. Fidelity is the exactness of a distorted image relative to its original image. Perception is inspired from characteristics of human visual systems HVS. This type of metric considers for example visual attention [2], contrast masking [3] etc. Aesthetics is subjective and may contain [1] visual constancy, visual attention and visual fatigue etc.
3. IMAGE QUALITY ASSESSMENT

Due to the continuous development in the multimedia technologies, images have become a great source of information. While transmitting an image over wired or wireless network channel its quality is degraded due to the noise/distortions introduced. The image on the receiver side is no longer the same as the one sent at sender side. IQA is hard as the algorithm needs to be reliable, fast, robust and should require less information from the reference image.

4. THREE TYPES OF IQA

IQA methods are categorized into the following three categories. Methods are categorized based on the amount of information needed from a reference/original image.

4.1. Full-Reference IQA

Full Reference QA algorithms have the original/reference image available at hand for the comparison with a distorted version of the same image. These kinds of methods are best suited for in-lab testing as in the real use cases the reference image is not available. Figure 1 is reproduced from [1].

![Figure 1: FR IQA Framework [1]](image)

4.2. No-Reference IQA

In NR QA methods the reference image is not available at all. QA algorithm must be able to evaluate the quality of a distorted image as that is the only information available at hand. Figure 2 is reproduced from [1].

![Figure 2: NR IQA Framework [1]](image)
4.3. Reduced-Reference IQA

RR QA methods serve as a compromise between NR and FR methods. These methods require partial information from the reference image in the form of features. That partial information is sent to the receiver side through an ancillary channel. The goal here is to reduce the features required. Figure 3 taken from [4] indicates the framework for reduced reference IQA.

5. BENCHMARK DATABASES FOR IQA

In this section we will discuss some standard databases available for the evaluation of IQA methods. In order to evaluate a newly proposed IQA metric with the ones that are currently existing, one needs to perform experiments on the publicly available databases for the purpose. The databases contain reference images for which various distorted images are produced by using different distortion types. Each distorted image has as associated subjective score. To evaluate a new metric you need to compute the objective score of each distorted image and compute correlations between the objective and subjective scores.

5.1. CSIQ Database

This database [5] contains 30 reference images where each image is introduced with 6 types of distortions at 4 to 5 different levels of distortions. The distortions types include JPEG compression, JPEG-2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring. The subjective ratings are given in Differential Mean Opinion Scores (DMOS). The database is freely downloadable for the experimentation.

5.2. TID2013 Database

This database [6] contains larger test images set than its earlier release TID2008. It contains 25 reference images, 24 distortion types and 5 levels for each distortion type, which makes 3000 test images in total. The images and associated MOS values are freely available for the download and research purposes.

5.3. Tampere Image Database 2008 (TID2008) Database

This database [7] contains 25 reference images and 17 types with 4 levels of distortions for each image, in total 1700 images and subjective ratings are reported in the form of Mean Opinion Scores (MOS). The database is available freely for investigation of the existing and new metrics.
5.4 Laboratory for Image and Video Engineering (LIVE) Database

This database [8], [9] contains 29 reference images, each image distorted with 5 distortion types, total of 799 distorted images. Subjective ratings are provided in the form of DMOS.

5.5 IVC Database

This database [10] contains 8 original images distorted at 3 different processing and 5 different compression rates. In total there are 120 distorted test images available.

5.6 Cornell A-57 Database

This database [11] is relatively a small database. It contains 3 reference images, 6 distortion types at 3 different distortion levels. This database is publicly available for research.

5.7 Wireless Imaging Quality (WIQ) Database

This database [12] contains 7 reference images and 80 distorted images.

6. STATE-OF-THE-ART APPROACHES

This section discusses some recent approaches that have been proposed for the three types of assessment methods.

Table 1: Full reference approaches

<table>
<thead>
<tr>
<th>Serial No</th>
<th>Year</th>
<th>Method</th>
<th>Database</th>
<th>Performance Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2011</td>
<td>In spatial domain [15]</td>
<td>Standard MATLAB Vegetable Image</td>
<td>SCC</td>
</tr>
<tr>
<td>4</td>
<td>2015</td>
<td>Binocular Visual Characteristic [16]</td>
<td>NBU, LIVE I, LIVE II, MICT and CML</td>
<td>PLCC, SRCC and RMSE</td>
</tr>
<tr>
<td>5</td>
<td>2010</td>
<td>Gabor Filters [17]</td>
<td>Their own database explained in the paper</td>
<td>PCC, RMSE MOS, SCC and Outlier Ratio</td>
</tr>
</tbody>
</table>

6.1. Full Reference Approaches

Table 1 contains the studies that have been reviewed for FR methods. In this paper [13], a novel framework for calculating Image Quality Metric (IQM) is given in discrete wavelet domain using Haar wavelet. The framework is applicable to map based such as SSIM that generate quality maps and nonmap based metrics for example PSNR as a final score, with reduced complexity and improved accuracy. The framework has been applied to various well-known assessment methods SSIM, VIF, PSNR and AD. The framework provides trade-off between accuracy and complexity, hence, could be efficiently applied in wavelet based image/video processing applications.

This paper [14] presents a machine learning expert based on SVM classification and regression, known as Machine Learning based IQM (MLIQM), shown in Figure 4. The feature is composed
of local quality features on which multi class SVM classifier is applied to classify test image into one of the five quality classes as recommended by ITU. After this SVM regression is applied to score the quality of the test image. The proposed technique is compared with four FR IQA techniques i.e. VSNR, PSNR, VIF and MS-SSIM, have been reported to perform better than them.

This paper [15] proposes new FR IQA metric for objective assessment in spatial domain. This metric is a combination of four factors, namely, sensitivity, edge performance index, contrast improvement and structural correlation. The proposed framework is shown in Figure 5. The proposed metric is reported to be more reliable than SSIM across different distortion types.

This paper [16] proposes a framework that gives IQA method for stereoscopic 3D images. The proposed methodology consists of the two phases; training and quality estimation. Binocular receptive field properties are learnt to simulate in simple and complex cells in cortex and to project their impact in quality estimation.

This paper [17] proposed FRIQA for the images distorted with local geometric distortions. The framework is shown in Figure 6. At first, the displacement field and a set of features by using Gabor filters from the image structure are extracted. Local quality scores are then calculated by evaluating the impact of displacement field on the features. Error pooling is performed to calculate the overall quality metric.
Table 2: Reduced reference approaches

<table>
<thead>
<tr>
<th>Serial No</th>
<th>Year</th>
<th>Method</th>
<th>Database</th>
<th>Performance Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2015</td>
<td>Deep Learning, Restricted Boltzmann Machine Similarity Measure RBMSim [18]</td>
<td>LIVE Multiply Distorted Image Quality and CSIQ</td>
<td>SROCC and Pearson Correlation Coefficient</td>
</tr>
<tr>
<td>5</td>
<td>2012</td>
<td>2012 Structural Similarity Estimation [22]</td>
<td>LIVE, Cornell, IVC, Toyama-MICT, TID 2008</td>
<td>CSIQ, PLCC, MAE, RMS, SRCC, KRCC</td>
</tr>
<tr>
<td>6</td>
<td>2014</td>
<td>Phase Information in Complex Wavelet Domain [23]</td>
<td>LIVE</td>
<td>Correlation Coefficient (Prediction Accuracy) and Rank-Order Correlation (Prediction Monotonicity)</td>
</tr>
<tr>
<td>7</td>
<td>2015</td>
<td>Entropy Differences in DCT Domain [24]</td>
<td>CSIQ, Toyama and LIVE</td>
<td>Pearson Linear and Spearman Rank-Order Correlation Coefficient (SROCC)</td>
</tr>
</tbody>
</table>

6.2. Reduced Reference Approaches

Table 2 contains the studies that have been reviewed for RR methods. Novel stochastic RR-IQA metric [18], automatic stochastic procedure capable of assessing the quality of distorted images, independent of the type of image of distortions. Compared with subjective IQA algorithms and two benchmark databases. Restricted Boltzmann Machines are stochastic neural networks and contain two layers of neurons, the hidden layer h and the visible layer v, used for unsupervised
learning. The learning algorithm used for BMs is called Contrastive Divergence CD. Framework is called Gaussian Bernoulli Restricted Boltzmann Machine (GRBM). Each pixel from the RI is modelled on the visible layer v, with individual RGB value encoded as one neuron. For easy weight manipulation the image was split into sub-images. The distorted image DI quality can be assessed by initializing the v of GRBM learned on the RI by the sub images from DI, this is called vDI. After this step, Gibbs sampling is performed to infer the values of hidden neuron and then values of visible neurons are reconstructed based on the hidden neuron values. The reconstructed values are called vDI*. The Root Mean Squared Error RMSE distance (or any other metric) between vDI and vDI* reflects the quality of distorted image. GRBM trained on the RI has to be sent to the client side and very less information is required from the reference image. The performance is comparable with widely known FR- IQA metrics and its computation time is less. So, it can be applied to video/multimedia applications to be able to do real-time performance evaluation of network services.

In [19] the quality score is estimated by measuring the changes in weighted entropies of reference and distorted images in the wavelet domain. Wavelet coefficients of natural image and distorted are modelled as GSM distributions. The quality metric is obtained as a distance between reference and natural image approximation of a distorted image. The proposed algorithm perform much better than FR MSE algorithm.

In this paper [20] the IQA metric is based on Discrete Cosine Transform DCT coefficients, the method follows three steps, shown in Figure 7. In the first step, block-based DCT is applied, as a result of which we have DC coefficients where the most of the energy of the image resides and AC coefficients which can be easily ignored because energy is very low within these coefficients. The DCT coefficients are reorganized by decomposing them into 10 sub bands. Then the coefficients of the same sub bands from different blocks are grouped together such that there is a structural similarity between all the sub bands. The reorganized DCT coefficients look like wavelet image representation. In the second step, generalized Gaussian density GGD curve fitting is done to model DCT coefficient distribution GGD model efficiently represent the coefficient histogram for each DCT subband. In addition to two standard parameters of GGD another parameter, prediction error is introduced. In the last step, city block distance is computed between actual distribution and the fitted distribution. The proposed method has been tested on LIVE database and compared with the well-known RR WNISM and FR PSNR and SSIM. The proposed method outperforms the RR WNISM and FR PSNR but slightly underperforms the FR SSIM.

In [21] a novel RR IQA algorithm based on image statistics is proposed. Image statistics are modelled in pixel domain which is based on gradient distribution. RR quality features are extracted based on statistical model. The quality is measured by determining the similarity between reference and distorted image. The proposed method is reported to perform as good as widely known FR PSNR.
In this paper [22], an approximation attempt for RR SSIM metric from FR SSIM has been made by divisible normalization transform DNT of the natural image. In addition to quality assessment, image repair concept has been introduced by matching the sub band statistical properties of the distorted image with those of the reference image. Features are extracted from the reference image by applying multi-scale linear transform (wavelet transform) and then DNT representation is calculated by dividing each coefficient of wavelet by a local energy measure based on neighbouring coefficients. For the efficient summarization of the statistical properties of the reference image zero mean GGD is fitted over DNT coefficients. For the estimation of RR SSIM, the effect of the distortions on the statistical properties need to be coherent with corresponding FR SSIM. The proposed method is compared with FR (PSNR and SSIM) and RR (Wavelet Marginal and DNT marginal) and the proposed metric is highly competitive in most of the cases.

In this paper, [23] IQA is based on the Complex Wavelet Transform (CWT) as it carries not only the magnitude information but also the phase information. Choosing the right RR features for IQA is critical, they have to be representative and yet small. Here different features are used namely, KLD, the standard deviation, the kurtosis, the skewness in the distribution of the relative phase and directional information from CWT coefficients. Generalized complex wavelet coefficient known as dual-tree complex wavelet transform (DT-CWT) has been used rather than simplistic CWT. The central idea of the algorithm is that the different distortion types affect the distribution in different ways. The distribution of phases of CWT coefficients is uniform and therefore contain no information. But the distribution of relative phases of different images follows a similar pattern when affected by a single distortion type. In case of different distortion type the pattern would be different. General Regression Neural Network GRNN has been used to predict the quality of distorted image by regression. The performance has been compared with 2 FR-IQA and 5 RR-IQA metrics and the proposed method is reported to be as good as popular FR-IQA metric MS-SSIM while using smaller features. To assess the quality of the distorted image we need the histograms of the relative phase of the distorted and the original image and then compute the Kullback-Liebler divergence.

In this paper [24], Figure 8, the differing importance of spectral sub-bands in DCT domain of images have been exploited by considering the fact that human eye corresponds differently to different bands. Quality of image should be assessed separately in each sub-band. Quality of the distorted image is computed via Shannon entropy differencing of DCT co-efficients between the reference and the distorted image. The overall quality of the distorted image is calculated by the weighted summation over entropy difference of all the sub bands. On the sender side the feature extraction is performed in three steps i.e. firstly, block-based DCT of the RI, secondly, the coefficient rearrangement and lastly the entropy calculation of each band. The same process is performed on the receiver side with the DI. The entropy features of RI at the sender side are sent to the receiver side over the ancillary error free channel which is required at the receiver side for the differencing. Higher prediction accuracy was attained using less reference data, similar frequency bands were merged into one term. The proposed algorithm was compared with already existing RRED, WNISM and RRVIF and overall outperforms over the three methods.
6.3. No Reference Approaches

Table 3 contains the studies that have been reviewed for NR methods. The proposed framework [25] is a combination of two existing NR-IQA approaches, shown in Figure 9. The two approaches are called free energy based distortion metric FEDM and structural degradation model SDM. The newly proposed metric no reference free energy and structural degradation model NFSDM is reported to outperform two FR metrics, SSIM and PSNR and NR metrics BLIINDS-II [26] and DIIVINE [27]. This metric makes use of a relatively novel feature which is a non-linear
combination of free energy and structural degradation information. This metric has been tested on LIVE database.

This paper [28] proposes shearlet based no reference image quality assessment SHANIA framework, Figure 10. The method makes use of natural scene statistics NSS model in shearlet domain [29]. The central idea behind approach is that the statistical properties of natural images remain constant in shearlet domain and change for the distorted images. The natural parts from the distorted image are used as reference and quality score is computed between reference and the distorted parts. The problem is mapped into a classification problem. The features are obtained from the distorted and reference parts and softmax classifier is used to obtain the quality score. The proposed framework is comparable in performance to three FR metrics including PSNR, SSIM and Multi-Scale SSIM and four NR methods BIQI, DIIVINE, BLINDS II and BRISQUE.

The paper [30] proposes NR-IQA framework based on visual codebooks which is a generalized framework not particular for some types of distortions, as shown in Figure 11. The proposed framework is reported to outperform FR-IQA SSIM and PSNR on LIVE database. The proposed method makes use of local feature extracted by using Gabor-filters. After this stage, visual codebook is constructed by forming a large set of block based Gabor feature vectors for the training images. Codebook C is constructed from that set by using clustering algorithm. Images
are represented in the form of codewords taken from codebook. Afterwards, regression is applied to calculate the quality score.

The proposed framework [31] is for multiple domain images in real time applications, shown in Figure 12. There are three steps involved in this method, (1) local feature extraction, (2) global feature extraction and (3) regression model. In this approach step 1 and 3 unlike previous approaches are not treated independently rather they are linked together by back projection. Back projection allows to choose less number of discriminative features. The method utilizes relatively small set of features obtained from raw image patches directly which makes it a fast method. It has been applied to the document image datasets to prove its cross-domain applicability.
The proposed framework [32] utilizes local spatial and spectral features of distorted images, overview in Figure 13. Two types of features have been used in the method, namely, spatial and spectral entropies. The method involves four stages, i.e., pre-processing(down-sampling) of the distorted image, block-based partition and computation of spectral and spatial entropies for each block, percentile pooling from the two feature sets is done and lastly the quality score is computed. The metric is called Spatial-Spectral Entropy based Quality index SSEQ and have been reported to outperform FR SSIM and NR DIVINE, BLINDS II and BIQI on LIVE and TID2008 database.

Figure 13: Framework for NR-IQA based on spatial and spectral entropies [32]

7. DISCUSSION

The problem of IQA is inherently hard. It has been dealt by using different tools. In this paper we have discussed entropy based methods, gradient based methods and DCT based methods. We have also shown a framework where deep learning has been used to deal with it. The fact that images from different domains have different properties. We represent those properties as features to a machine. Therefore, we cannot have universal set of features that could be applicable to all types of applications. State-of-the-art methods for the three types of IQA have been discussed previously. It shows that we need new IQA metrics that possess greater power than widely used PSNR and MSE.

8. CONCLUSIONS

We have reviewed some state-of-the-art approaches for the three types of IQA. Of the three types, NR IQA poses most challenges. As it requires no information from the reference image at the same time without strict assumptions regarding distortion level and types. Till date no system have been developed that meets the requirements of real-time quality prediction, data-base independence and generalized for all the distortion types and levels.

ACKNOWLEDGEMENTS

This publication was made possible by NPRP grant # NPRP8-140-2-065 from Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.
REFERENCES


