New Adaptive Image Quality Assessment Based on Distortion Classification

1 Xin JIN, 1 Mei YU, 2 Gangyi JIANG, 1 Feng SHAO, 1 Fen CHEN, 1 Zongju PENG
1 Faculty of Information Science and Engineering, Ningbo University, 315211, China
2 National Key Lab of Software New Technology, Nanjing University, Nanjing 210093, China
Tel.: +86-574-87600017
E-mail: Jin_Xin1006@163.com

Received: 14 October 2013 / Accepted: 9 January 2014 / Published: 31 January 2014

Abstract: This paper proposes a new adaptive image quality assessment (AIQA) method, which is based on distortion classifying. AIQA contains two parts, distortion classification and image quality assessment. Firstly, we analysis characteristics of the original and distorted images, including the distribution of wavelet coefficient, the ratio of edge energy and inner energy of the differential image block, we divide distorted images into White Noise distortion, JPEG compression distortion and fuzzy distortion. To evaluate the quality of first two type distortion images, we use pixel based structure similarity metric and DCT based structural similarity metric respectively. For those blurriness pictures, we present a new wavelet-based structure similarity algorithm. According to the experimental results, AIQA takes the advantages of different structural similarity metrics, and it’s able to simulate the human visual perception effectively.

Keywords: Image quality assessment, Structural similarity, Wavelet based structural similarity, Distortion analysis, Adaptive metric selection.

1. Introduction

Image acts as vital source of information to human beings, whose quality will play a decisive implication on the accuracy and integrity of acquiring information. Thus, assessment of an image becomes a crucial task in image processing application [1]. The goal of image quality assessment (IQA) is to calculate the extent of quality degradation and is thus used to evaluate the performance of processing systems and optimize the choice of parameters in processing [2].

The human visual system (HVS) is the ultimate receiver of the majority of processed images, and evaluation based on subjective experiments is the most reliable way of IQA. However, it has some critical drawbacks that a large number of images and tests are needed. Therefore, the objective IQA is preferred in practical situations and thus has been widely investigated [3].

The earliest and most widely used methods for the quality evaluation are MSE (mean square error) and PSNR (peak signal-to-noise ratio). They are both based on pixel-level evaluation algorithms and lack of considering the relationship between pixels. Thus, the evaluation results differentiate greatly from the perception of HVS [4]. Since cluster of natural images occupies an extremely tiny portion in the image space and they are highly structured with samples having strong neighbor dependencies. Also HVS is an information extractor that seeks to identify...
and recognize objects, highly sensitive to the structural distortions and automatically compensates for the non-structural distortions. In order to compensate the disadvantages of MSE/PSNR and incorporating the importance of structural information SSIM (structural similarity) metric is introduced [5]. It is made up of easy to compute statistics (mean, variance and covariance of small patches) of luminance comparison, contrast comparison and structural comparison. The evaluation result of SSIM has a relatively high consistency with HVS, especially in the case of assessing White Noise polluted images. Based on the thought of SSIM metric, a number of quality metrics are created. Reference [6] introduces a new SSIM method in frequency domain based on DCT transform. It replaces luminance structural similarity with the DCT coefficients structural similarity, and achieves a better result of evaluating the JPEG compression distorted images. Reference [7] proposes a feature similarity (FSIM) metric based on the fact that HVS understands an image mainly according to its low-level features. Considering the phase congruency (PC) is contrast invariant while the contrast information does affect HVS’ perception of image quality, the image gradient magnitude is employed. In general, these metrics process images from the perspective of the whole image equally, rather than separating the interested regions for viewers. Since there exist different degree of attention in edge, texture and flat region, reference [8] presents a new method using image edge combined structure similarity. It uses the smoothing function to detect the edge regions in different wavelet scales. Based on the detection, an edge-similarity map is accomplished, and finally a similarity map is obtained with the combining of edge-similarity map and luminance-similarity map. In reference [9], the edge regions are detected in pixel domain firstly, and a high-pass filter is used to filter the edge regions aiming to strengthen the edge information secondly. Finally the enhanced edge similarity method is formulated by adopting SSIM method. However, all the models mentioned above, no matter using global processing or local reinforcement, measure different kinds of distortion images by means of same metric. The short coming existed is that it limits the ability of assessing specific distortion images. Reference [10] proposes a metric which extracts and fusions different features for different distortion images by machine learning method. Since the distortion type of images is not confirmed before evaluating them, this approach is only applicable to the known type of distortion images in quality assessment.

In order to effectively avoid the problems mentioned above, we propose an adaptive IQA (AIQA) from the perspective of distortion classification. Through the process of distortion classification, we utilize the appropriate metric to assess different distortion type images. The experimental result shows that AIQA takes advantages of different quality metrics, and have a high consistency with the subjective perception of human eyes for all kinds of distortion types.

2. Adaptive Image Quality Assessment Based on Distortion Classification

View from the image distortion characteristics, noise, blurring and blockiness are the most common features to distorted images [11]. However, objective IQA metric usually has different evaluation accuracy due to the different type of distortion. From these two perspectives, this paper presents an adaptive evaluation algorithm, which contains two parts. In the first part, analyze the characteristics of different distortion types. And based on these features, propose a distortion classification method. Besides, the distorted images are divided into three categories, namely White Noise (WN) distortion, JPEG distortion and fuzzy distortion, here fuzzy distortion includes Gaussian blur (Gblur) distortion, JPEG2000 (JP2K) distortion and fast fading (FF) distortion. In the second part, choose appropriate quality metric to evaluate distorted image according to its distortion type. Fig. 1 is the theoretical diagram.

2.1 Distortion Classification

When an image (Fig. 2(a)) suffers from WN pollution (Fig. 2(b)), the diagonal coefficients (DC) of wavelet sub-band obey the generalized Gaussian distribution approximately [12] as Fig. 2(d) showed. In Fig. 2(c) and (d), the horizontal axis stands for diagonal coefficients, and the vertical axis is the frequent. For JPEG distorted image, a block artifact will occur in the pixel domain, that is to say, a luminance jump happens to the edge of coding block. Therefore, the amplitude of line differential luminance (or column differential luminance) at the edge of block will be higher than that in the inner of block. Based on these two features, we have designed a distortion classification method, and the specific steps are described as below.

i). Segment original image \(X\) and distorted image \(Y\) into size of \(64 \times 64\) image blocks (non-overlapping), and get the image block as \(x_i, y_i\) respectively, where \(i = 1, 2, \cdots, M, M\) stands for the total number of image blocks. Then
transform $x_i, y_j$ into wavelet domain using Haar wavelet. Estimate the standard deviation of noise in each block by formula (1) [13]:

$$\sigma_n = 1.4826A_n^H,$$

$$\sigma_n = 1.4826A_n^H,$$  \hspace{1cm} (1)

where $A_n^H, A_n^H$ represents the median amplitude of diagonal coefficients of $x_i, y_j$ in wavelet domain, respectively. Then calculate the distance of standard deviation between each block

$$\Delta\sigma = \sigma_n - \sigma_n.$$  \hspace{1cm} (2)

If $\Delta\sigma > Th_{WN}$, where $Th_{WN}$ is a threshold, $Y$ will be confirmed as a WN distorted image. Otherwise, go to step ii.

ii). Compute the differencing luminance along each horizontal line of $X$ and $Y$

$$X^h(i, j) = |X(i, j) - X(i, j + 1)|,$$  \hspace{1cm} (4)

where $m, n$ stands for the height and width of $X$; $X(i, j)$, $Y(i, j)$ are the luminance value of $X$ at the point of $(i, j)$, $i = 1, 2, \cdots, m$; $j = 1, 2, \cdots, n - 1$. $Y^h$ can be obtained in the same way. Then segment $X^h$ and $Y^h$ into $8 \times 8$ blocks (non-overlapping). Thus we get the differential blocks as $x_i^h$ and $y_j^h$, where $i = 1, 2, \cdots, M^h, M^h$ is the total number of image blocks of $X^h$ or $Y^h$. Here we define the inner-energy and edge-energy as

$$Ex_i^h = \frac{1}{M^h} \sum_{p=1}^{M^h} \sum_{q=1}^{M^h} x_i^h(p, q),$$  \hspace{1cm} (5)

$$Ex_i^h = \frac{1}{M^h} \sum_{p=1}^{M^h} \sum_{q=1}^{M^h} x_i^h(p, q),$$

where $x_i^h(p, q)$ stands for the coefficient of $X^h$ at the point of $(p, q)$. With the same method, we can get $Ey_i^h$ and $Ey_i^h$. Then compute the ratio of edge-energy and inner-energy of each block

$$P_i = Ex_i^h(Ex_i^h)^{-1},$$  \hspace{1cm} (6)

$$P_i = Ey_i^h(Ey_i^h)^{-1}.$$  \hspace{1cm} (6)

Get the number of blocks which satisfy the inequality

$$P_i > P_i^*.$$  \hspace{1cm} (7)

Thus we get the judge criterion

$$J = N_i(M^h)^{-1}.$$  \hspace{1cm} (8)

If $J > Th_{JPEG}$, where $Th_{JPEG}$ is a threshold, $Y$ will be confirmed as a JPEG distorted image. Otherwise, go to step iii.

iii). Confirm $Y$ as a fuzzy image.

2.2. Pixel Based Structure Similarity

As in [5], the local SSIM index is defined as

$$SSIM(x, y) = \left[ l(x, y) \right]^{\alpha} \left[ c(x, y) \right]^{\beta} \left[ s(x, y) \right]^{\gamma},$$  \hspace{1cm} (9)

where $x, y$ is the image block; $\alpha, \beta$ and $\gamma$ are the weight adjustment factors; $l(x, y), c(x, y), s(x, y)$ is the luminance similarity, contrast similarity and structure similarity respectively.

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3},$$  \hspace{1cm} (10)

where $\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}$ are the average and standard deviation of $x$ and $y$; $\sigma_{xy}$ is the covariance of $x$ and $y$; $C_1, C_2$ and $C_3$ are positive constants used to avoid instability.

The similarity between $X$ and $Y$ is computed by

$$Q = SSIM(X, Y) = \frac{1}{N} \sum_{i=1}^{N} SSIM(x_i, y_i),$$  \hspace{1cm} (11)

where $N$ is the total number of image blocks.
2.3. Frequency Based Structure Similarity

According to [6], the local FSSIM index is

\[ \text{FSSIM}(x, y) = \left[ l(x, y) \right]^\alpha \left[ c(x, y) \right]^\beta \left[ f(x, y) \right]^\gamma, \]  \tag{12}

where \( l(x, y), c(x, y), \alpha, \beta \text{ and } \gamma \) have the same meaning with those in SSIM metric, and \( f(x, y) \) stands for the frequency based structure similarity

\[ f(x, y) = \frac{\sigma_{F_x} + C_3}{\sigma_{F_x}^2 + \sigma_{F_y}^2 + C_3}, \]  \tag{13}

where \( \sigma_{F_x} \) and \( \sigma_{F_y} \) are the standard deviation of \( x \) and \( y \) in DCT domain; \( \sigma_{F_{xy}} \) is the covariance of \( x \) and \( y \) in DCT domain. They can be computed as

\[ \begin{align*}
\sigma_{F_x}^2 &= \sum_{u,v} w_{uv} [F_{xuv}(u,v) - \mu_{F_x}]^2 \\
\sigma_{F_y}^2 &= \sum_{u,v} w_{uv} [F_{yuv}(u,v) - \mu_{F_y}]^2 \\
\sigma_{F_{xy}} &= \sum_{u,v} w_{uv} [F_{xuv}(u,v) - \mu_{F_x}][F_{yuv}(u,v) - \mu_{F_y}],
\end{align*} \]  \tag{14}

where \( F_{xuv}(u,v) \) and \( F_{yuv}(u,v) \) are AC coefficients in DCT domain of image block \( x \) and \( y \); \( \mu_{F_x} \) and \( \mu_{F_y} \) are the weighted expectation, and \( w_{uv} \) is the weighting factor which is normalized according to the quantization parameter of DCT coefficients in JPEG compression [14].

Similarly, the overall similarity in DCT domain is obtained by averaging the local FSSIM index

\[ Q = \text{FSSIM}(X, Y) = N^{-1} \sum_{i=1}^{N} \text{FSSIM}(x_i, y_i). \]  \tag{15}

2.4. Wavelet Based Structure Similarity

According to the special characteristic of HVS and the contrast sensitivity function (CSF), the median frequency regions draw the dominant attention when viewing an image. And the information of this part will be reflected in the appropriate component of wavelet transform. Therefore, we propose a wavelet based structure similarity (WSSIM) metric to evaluate fuzzy images. The details of WSSIM are presented as below.

Firstly, transform \( X \) and \( Y \) into wavelet domain, and extract the appropriate component as \( X_s \) and \( Y_s \), respectively. Apply SSIM to \( X_s \) and \( Y_s \), get the similarity between original and distorted image blocks using a sliding window which runs across the image from up-left to bottom-right. The local WSSIM index is given by

\[ \text{WSSIM}(s', y') = \left[ l'(s', y') \right]^\alpha \left[ c'(s', y') \right]^\beta \left[ f'(s', y') \right]^\gamma \]  \tag{16}

\[ l'(s', y') = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c'(s', y') = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s'(s', y') = \frac{\sigma_{s', y'} + C_3}{\sigma_{s'} + \sigma_{y'} + C_3}, \]  \tag{17}

where \( x' \) and \( y' \) stands for the image block of \( X_s \) and \( Y_s \); \( l'(s', y') \), \( c'(s', y') \) and \( s'(s', y') \) is the sub-band coefficient similarity, contrast similarity and structure similarity, respectively; \( \mu_x \), \( \mu_y \), \( \sigma_x \) and \( \sigma_y \) is the average and standard deviance of \( x' \) and \( y' \); \( \sigma_{s', y'} \) is the covariance of \( x' \) and \( y' \).

The similarity of the whole image is obtained by averaging the local WSSIM index as well

\[ Q = \text{WSSIM}(X, Y) = N^{-1} \sum_{i=1}^{N} \text{WSSIM}(x_i', y_i'). \]  \tag{18}

where \( N \) is the total number of image blocks.

3. Experiment Results

The image data we used comes from LIVE database. In LIVE database, there are 29 original images and 779 images distorted by WN, JPEG, Gblur, JP2K and FF with the differential mean opinion score (DMOS) [15].

3.1. Distortion Classification

In the distortion classification part, there are two parameters need to be confirmed: \( Th_{WN} \) and \( Th_{JPEG} \). To obtain these two parameters, we divided the database into training set and testing set. In the training set, there are 12 original images and the corresponding distorted images, including 60 WN, Gblur and FF images, 70 JPEG and JP2K images. Besides the texture complexity degree of the originals contains being low, median and high. Images remained are sorted into testing set, including 17 originals and 459 distorted images: 85 WN, Gblur and FF images, 105 JPEG images, 99 JP2K images.

Step 1. Calculate the Gσ (equation (3)) of each distorted image. Then in the range of \([-0.5, 1.5]\), take a number every 0.05 as the testing value of \( Th_{WN} \). For each testing value, compute the accuracy of distortion classification, and plot the curve of accuracy and testing value as Fig. 4(a). Fig. 4(a) manifests that, for WN images, the accuracy decrease from 1 when \( Th_{WN} > 1.3 \). And for non-WN images, the accuracy increase to 1 when \( Th_{WN} > 0.45 \). Thus, we can use \( Th_{WN} = 0.8 \) to tell WN images apart from other distorted images.

Step 2. Calculate the ratio of edge-energy and inner-energy of each distorted image, exclude the WN ones, and get the classification criteria \( J \). Then take a number every 0.01 as the testing value of
\( T_{\text{JPEG}} \) in the range of \([0.4, 0.7]\). For each \( T_{\text{JPEG}} \), compute the accuracy of distortion classification and plot the curve as well. The experimental result can be seen in Fig. 4(b). Fig. 4(b) tells that the classification algorithm starts to misjudge at the point of \( T_{\text{JPEG}} = 0.65 \). And for those non-JPEG images, they all will ascend to 1 if \( T_{\text{JPEG}} > 0.53 \). Therefore, we choose 0.57 as the value of \( T_{\text{JPEG}} \) in this work.

![Fig. 4. Threshold test.](image)

With the thresholds, we examine the performance of classification accuracy on the testing set. Results are presented in Table 1, where the first line means the true distortion type and the first column means the classified distortion type. Besides, except the bottom line, the data in Table 1 represent the number of truly classified images for the corresponding type of distortion. From the accuracy line, we can conclude that the distortion classification method can classify the distorted images without mistake.

### 3.2. Quality Assessment

In this part, the whole database is involved, and we set the adjustment factor as \( \alpha = \beta = \gamma = 1 \), two thresholds as \( T_{\text{WN}} = 0.8 \), \( T_{\text{JPEG}} = 0.57 \). Besides the objective quality metric is evaluated by comparing the DMOS and the DMOS, which is predicted by \( Q \) and DMOS through logistic-4 regression function, using four performance metrics, and they are linear correlation coefficient (CC), Spearman rank order correlation coefficient (SROCC), out ration (OR) and root mean square error (RMSE). The results are presented in Table 2.

### Table 1. Accuracy of distortion classification algorithm.

<table>
<thead>
<tr>
<th>Classified Type</th>
<th>True Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>WN</td>
<td>JPEG</td>
</tr>
<tr>
<td>WN</td>
<td>85</td>
</tr>
<tr>
<td>JPEG</td>
<td>0</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 shows that the AIQA metric has take the advantage of SSIM, FSSIM and WSSIM very well. For each distortion type, the assessment result of AIQA is equal to the best result of those three metrics. Besides, Fig. 5 shows the logistic regression result of proposed algorithm. To faithfully reflect the DMOS, scatter plots should be close to the dotted line. It’s easy to know that the proposed algorithm fitted the dotted line well. In other words, the proposed quality metric well reflects the DMOS.

### 4. Conclusions

With the rapid development of image processing, 2D image quality assessment (IQA) has achieved significant breakthrough. However, pursuing the accuracy of IQA, the computational complexity needs to be considered more and more urgently. In this paper, an adaptive IQA (AIQA) is proposed from the perspective of distortion classification. We take the simple and practical strategy of assessing the corresponding classified type of distortion, which can be done separately. The experimental result shows the objective quality is highly consistent with the perceptual quality of human visual system (HVS). What’s more, the biggest innovative point lies in putting forward a judgmental approach for image distortion classification theoretically. It’s flexible and wide-spread application. Whereas, the proposed approach still has a shortcoming: fail to further improve the function to separate the following types of distortion: Gaussian blur, JPEG2000 compression and fast fading. The future work will focus on settling this problem and taking the distortion classification metric into the stereoscopic IQA.
Table 2. Performance of different metrics.

<table>
<thead>
<tr>
<th>Distortion</th>
<th>CC</th>
<th>SROCC</th>
<th>OR</th>
<th>RMSE</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>WN</td>
<td>0.9701</td>
<td>0.9649</td>
<td>0.2345</td>
<td>3.8773</td>
<td>SSIM [5]</td>
</tr>
<tr>
<td></td>
<td>0.9503</td>
<td>0.9462</td>
<td>0.3931</td>
<td>4.9717</td>
<td>FSSIM [6]</td>
</tr>
<tr>
<td></td>
<td>0.9590</td>
<td>0.9431</td>
<td>0.3172</td>
<td>4.5265</td>
<td>WSSIM</td>
</tr>
<tr>
<td>JPEG</td>
<td>0.9307</td>
<td>0.9061</td>
<td>0.2743</td>
<td>5.8480</td>
<td>SSIM [5]</td>
</tr>
<tr>
<td></td>
<td>0.9426</td>
<td>0.9140</td>
<td>0.2286</td>
<td>5.3393</td>
<td>FSSIM [6]</td>
</tr>
<tr>
<td></td>
<td>0.9383</td>
<td>0.9054</td>
<td>0.2800</td>
<td>5.5303</td>
<td>WSSIM</td>
</tr>
<tr>
<td>Gblur</td>
<td>0.9426</td>
<td>0.9140</td>
<td>0.2286</td>
<td>5.3393</td>
<td>AIQA</td>
</tr>
<tr>
<td></td>
<td>0.9010</td>
<td>0.9143</td>
<td>0.5103</td>
<td>6.8220</td>
<td>SSIM [5]</td>
</tr>
<tr>
<td></td>
<td>0.9604</td>
<td>0.9620</td>
<td>0.4069</td>
<td>4.3800</td>
<td>WSSIM</td>
</tr>
<tr>
<td></td>
<td>0.9455</td>
<td>0.9340</td>
<td>0.3373</td>
<td>5.2751</td>
<td>AIQA</td>
</tr>
<tr>
<td>JP2K</td>
<td>0.9524</td>
<td>0.9495</td>
<td>0.3550</td>
<td>4.9406</td>
<td>SSIM [5]</td>
</tr>
<tr>
<td></td>
<td>0.9560</td>
<td>0.9519</td>
<td>0.3550</td>
<td>4.7526</td>
<td>WSSIM</td>
</tr>
<tr>
<td></td>
<td>0.9560</td>
<td>0.9519</td>
<td>0.3550</td>
<td>4.7526</td>
<td>AIQA</td>
</tr>
<tr>
<td>FF</td>
<td>0.9488</td>
<td>0.9490</td>
<td>0.2830</td>
<td>5.1963</td>
<td>SSIM [5]</td>
</tr>
<tr>
<td></td>
<td>0.9472</td>
<td>0.9420</td>
<td>0.2552</td>
<td>5.2726</td>
<td>FSSIM [6]</td>
</tr>
<tr>
<td></td>
<td>0.9533</td>
<td>0.9581</td>
<td>0.2276</td>
<td>4.9684</td>
<td>WSSIM</td>
</tr>
<tr>
<td></td>
<td>0.9533</td>
<td>0.9581</td>
<td>0.2276</td>
<td>4.9684</td>
<td>AIQA</td>
</tr>
</tbody>
</table>

Fig. 5. Scatter plots of $DMOS_p$ and $DMOS$. 
Acknowledgements

The research work was supported by National Natural Science Foundation of China under Grant Nos. 61271270, 61171163, 61271021.

References