A REFERENCE FREE QUALITY METRIC FOR COMPRESSED IMAGES

Rémi Barland, Abdelhakim Saadane
IRCCyN-IVC, UMR n°6597 CNRS, Ecole Polytechnique de l’Université de Nantes
Rue Christian Pauc, La Chantrerie, BP50609, 44306 Nantes, France.
Mail: {remi.barland, abdelhakim.saadane}@univ-nantes.fr

ABSTRACT
At high compression ratios, JPEG and JPEG-2000 compression techniques introduce distortions that are typically exploited by the No Reference assessment. In this paper, we propose to use these distortions (Blocking, Blurring and Ringing) to design a new reference free quality metric. The blocking measure only depends on the spatial information contained in the processed image, while the blurring and ringing measures moreover exploits the information of an importance map generated by the Osberger’s attention model. To produce the final quality score, a new pooling model is also proposed and discussed. The proposed metric is tested using two databases: the first one is composed of JPEG compressed images and the second one, of JPEG-2000 compressed images. Both of these comparative studies demonstrate the efficiency of the proposed quality metric.

1. INTRODUCTION
Subjective experiments represent the reference method for assessing image quality and involve the use of human subjects. The recommendation of the International Telecommunication Union (ITU) [1] is a commonly used standard for the subjective assessment of still images and video: it specifies the conditions of observation, the choice of observers, the material to use, etc. Even if they represent the reference assessment, subjective tests are costly, time consuming, difficult to practice and only valid for the particular set of viewing conditions that the test was performed under.

An alternative is the quality metrics. Most of the proposed image quality assessment approaches require the original image as a reference. The most widely used objective image quality metrics are certainly the Peak Signal-to-Noise Ratio (PSNR) and the Mean Squared Error (MSE), but they do not have a good correlation with the Mean Opinion Scores (MOS) resulting from subjective tests. To improve this correlation, a particular effort has been made to develop new objective image/video quality metrics that incorporate the Human Visual System (HVS) properties [2-5]. However, human observers can easily assess the quality of a distorted image without comparing to the original image. Hence, a blindly quality assessment is possible: such No-Reference (NR) approach seeks to assign quality scores that are consistent with human perception using a prior knowledge about the types of image impairments. For JPEG compressed images, at low bit rates, the two most annoying visual impairments are the blocking and the blurring effects, while in the case of JPEG-2000 compressed images, the blurring and ringing artefacts are the two most generated perceptual distortions.

In the literature, many metrics detecting and measuring blocking effect are presented. Gao et al [6] present a metric based on differences of averages of eight-pixel rows (columns) across vertical (horizontal) block edges. In the Phase Correlation method [7], the metric is based on a ratio using the inter-block and intra-block similarities. Some blocking metrics incorporate some properties issued from HVS processing. In the generalized block impairment metric (GBIM), described in [8], HVS masking is incorporated by means of weights derived from localized averages and standard deviations across block boundaries. In [9], the power spectrum of the 1-D absolute difference signal is computed using FFT. Luminance and texture masking are exploited to scale the signal before the frequency analysis.

Few blurring metrics exist: in [10], the blurring measure is computed using an histogram of DCT coefficients gathered from the compressed bit stream. Marziliano et al [11] compute the spatial extent of each edge in an image using inflection points in the luminance to localize the start and end of the edge. An other approach [12] consists to measure sharpness, the inverse of blurring, by calculating local edge kurtosis.

In [13], by viewing all edge points in JPEG-2000 compressed images as “distorted” or “un-distorted”, the authors propose, using principal component analysis, to
extract the local feature of a given edge point, which indicates both blurring and ringing. The local distortion metric is modeled using the probabilities of the given edge point being “distorted” or “un-distorted”. In [14], the algorithm uses a statistical model for wavelet coefficients and computes features that exploit the fact that quantization produces more zero coefficients than expected for natural images. The algorithms [13, 14] are based on a statistical model, which does not take into account the visual aspect of ringing and blurring distortions.

In this paper, the proposed method is to design separate models specifically tuned to certain type of distortion (blocking, blurring, ringing), and combine their results according to the impact of each type of impairment on the image quality. The blocking measure only depends on the spatial information contained in the processed image, while the blurring and ringing measures moreover exploits the information of an importance map generated by the Osherber’s attention model [15]. A general pooling model is used in order to predict quality score of compressed images. The performance of the proposed metric is evaluated using selected JPEG and JPEG-2000 compressed images in terms of correlation coefficients with regard to subjective test data.

2. NO REFERENCE QUALITY METRIC

At low bit rate, lossy coding potentially introduces visual impairments such as the blocking, blurring and ringing effects. Taking into account these observations, the proposed image quality metric (figure 1) is designed as follows: after a color conversion, the three distinct distortion measures of blocking (BIM), blurring (BM) and ringing (RM) effects are computed. At the last stage, a quality score is computed, using an inter-distortion pooling.

![Figure 1. Block diagram of the proposed NR Quality metric.](image)

2.1. Color space conversion

Digital images are usually coded in RGB color space, which cannot be used directly within a vision model, because the perception of color differences is highly non uniform. So, the first stage of the metric performs the conversion from RGB color space of the JPEG compressed image to an opponent color space. This color space is inspired by the processing taking place in the HVS, which presumably uses separate channels for black-white, red-green and blue-yellow information.

An image presented to the retina is sampled by cone photoreceptors maximally sensitive to Long (L cones), Middle (M cones), or Short (S cones) wavelengths. The RGB values are converted in LMS values by using transformations given in [16, 17]. Several opponent color spaces exist. In our lab, the Krauskopf color space [18] has been validated [19]. In this space, the decomposition into three components, called A, Cr1 and Cr2, corresponds to the three separate channels of the HVS. Only the achromatic component (A) is considered in this paper.

2.2. The No Reference blocking artefact measure

The blocking artefacts visually create an artificial discontinuity between neighboring blocks in a JPEG compressed image, caused by a severe and independent quantization of DCT coefficients of each block. This impairment can be amplified by the contrast between the neighbouring blocks. Considering these facts, the local blocking measure \( LBIM(k,l) \) for a block \((k,l)\) can be formulated as follows:

\[
LBIM(k,l) = \frac{R_{g}(k,l) + R_{l}(k,l)}{2} \cdot S(k,l)
\]

where \( R_{g}(k,l) \) (respectively \( R_{l}(k,l) \)) defines the contrast reinforcement produced by the horizontal (respectively vertical) neighboring blocks and \( S(k,l) \), the quantization severity.

For the block \((k,l)\), the quantization severity is defined as:

\[
S(k,l) = \frac{1}{1 + \text{STD}(k,l,a)}
\]

Where \( \text{STD}(k,l) \) is the standard deviation of the block \((k,l)\), and \( a \), a constant. The horizontal contrast reinforcement is computed as follows:

\[
R_{h}(k,l) = 1 + C_{h}(k,l)
\]

with

\[
C_{h}(k,l) = \frac{\max[A(k,l) - A(k,l-1), A(k,l) - A(k,l+1)]}{2\max[A(k,l) - A(k,l-1), A(k,l) - A(k,l+1)]}
\]

where \( A(k,l) \) (respectively \( A(k,l-1) \) and \( A(k,l+1) \)) is the mean value of the achromatic component for the block \((k,l)\) (respectively block \((k,l-1)\) and block \((k,l+1)\)). The vertical contrast reinforcement is defined with the same formula, but considering the vertical neighboring blocks. Here, we assume that the sensitivity of HVS to horizontal and vertical blocking artefacts are similar.
The final blocking measure $BIM$ is obtained using a Minkowski summation:

$$ BIM = \left( \frac{1}{N_{B_V}N_{B_H}} \sum_{k=1}^{N_{B_V}} \sum_{l=1}^{N_{B_H}} (LBIM (k,l))^p \right)^{1/p} $$

where $N_{B_V}$ (respectively $N_{B_H}$) represents the number of vertical (respectively horizontal) blocks in the processed image.

2.3. The No Reference blurring and ringing artefact measures

Blur in an image is due to the attenuation of the high spatial frequency coefficients, which occurs during the compression stage. Visually, the blurring effects correspond to a total distortion on the whole image, characterized by an increase of the spread of edges and spatial details, while the ringing effect locally produces haloes and/or rings near sharp object edges in the frame. Furthermore, the HVS does not carry out a systematic and local research of these impairments in the whole image, but rather, it identifies some regions of interest for judging the perceptual quality. Considering each of these observations, we proposed in [20], two distinct measures of blurring and ringing distortions, based on the spatial information contained in both of an importance map generated by Osberger’s attention model [15] on the processed image.

Considering this fact, the formula of the blurring measure (BM) is formulated with a ratio using spatial information and pixel activity and weighting by the Importance Map (IM) values:

$$ BM = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} IM (i,j)A'_{edge} (i,j)I'_3 (i,j)N (A'_{edge})}{M \times N} $$

Where $I_A(i,j)$ is the pixel $(i,j)$ intensity of $A$ component of size MxN pixels. $A_{edge}$ is the binary image resulting from A edge detection. $A'_{edge}$ is the $A_{edge}$ complementary binary image. $N(A_{edge})$ (respectively $N(A'_{edge})$) is the number of non-null pixel values of $A_{edge}$ (respectively $A'_{edge}$).

Before measuring the ringing distortion, the areas around edges, called “ringing regions”, must be identified. These are computed by using a binary “ringing mask” on the current image, resulting from the detection and the dilation of strong edges. The “ringing region” image $(A_{ringing Mask})$ is computed by using a binary “ringing mask” on the current image. Then, a measure of ringing artefact is computed, defined by the ratio of regions activities of middle low and middle high frequencies, localized in these “ringing regions”. Each part of the defined ratio is locally weighted by the Importance Map values:

$$ RM = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} IM (i,j)A'_{RM edge} (i,j)I'_3 (i,j)N (A'_{RM edge})}{N (ringing Mask)} $$

Where $I_{RM}(i,j)$ is the pixel $(i,j)$ intensity of $A_{RM edge}$ image of size MxN pixels. $A_{RM edge}$ is the binary image resulting from $A_{ringing Mask}$ image edge detection. $A'_{RM edge}$ is the combination (XOR operator) of $A_{RM edge}$ complementary binary image and $ringing Mask$ binary image. $N(A_{RM edge})$ (respectively $N(A'_{RM edge})$ or $N(ringing Mask)$) is the number of non-null pixel values of $A_{RM edge}$ (respectively $A'_{RM edge}$ or $ringing Mask$) binary image.

2.4. The inter-distortion pooling

The proposed inter-distortion pooling model includes two parts. The first one is formulated as a linear combination of the first order terms of the blurring, blocking and ringing measures. The second part introduces the different associated crossed terms. This is expressed as follows:

$$ pMOS = a_0 + a_1 BIM + a_2 RM + a_3 BIM.BM + a_4 RM.BM + a_5 BIM.RM $$

Where $a_i$, $i=0..6$ are the weights to be optimised. Many formulations of the pooling model have been tested.

3. EXPERIMENTS AND RESULTS

3.1. Experimental validation of the blocking measure

The image database we use in the experiments is from [21]. It consists of 29 original high-resolution 24-bits/pixel RGB color images (typically 768x512) and their JPEG compressed versions with different compressed ratios. The total number of images in the database is 227. The bit rates used for compression are in the range of 0.03 to 3.2 bits per pixel, chosen such that the resulting distribution of quality scores is roughly uniform over the entire range. About 25 human observers assess the quality of each image as “Bad”, “Poor”, “Fair”, and “Good” and “Excellent”. The raw scores for each subject are normalized by the mean and the variance of that subject and then scaled and shifted by the mean and the variance of the entire subject pool to the full range (1 to 100).
Mean scores are then computed for each image after removing outliers.

Currently, the Video Quality Expert Group (VQEG) considers the standardization of quality assessment methods as one of its working directions. In order to quantify the performance of a quality metric, some statistical tools [22] are proposed: the Pearson linear correlation for the accuracy, the Spearman rank order correlation for the monotonicity, the outliers percentage for the consistency and the Kappa coefficient for the agreement. The Table 1 illustrates a performance comparison between the Wang’s blocking measure [9] and the proposed one.

<table>
<thead>
<tr>
<th></th>
<th>Pearson Correlation</th>
<th>Spearman Correlation</th>
<th>Outlier Ratio</th>
<th>Kappa Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>0.9518</td>
<td>0.9350</td>
<td>4.292%</td>
<td>0.71916</td>
</tr>
<tr>
<td>Blocking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>measure</td>
<td>Wang’s</td>
<td>0.9228</td>
<td>0.90175</td>
<td>7.296%</td>
</tr>
<tr>
<td>Blocking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Comparison of correlation results between Wang’s blocking metric and the proposed blocking measure.

The correlation results issued from the separate use of each blocking distortion measure allows to demonstrate the efficiency of the proposed blocking measure: the Pearson correlation obtained by the proposed blocking measure indicates a better ability to predict subjective scores with a minimum average error than the Wang’s blocking measure. The monotonic relationship (Spearman rank-order correlation) is respected. The smaller outlier ratio obtained by the proposed approach means that the proposed distortion metric has a better ability to provide consistently accurate predictions for all types of compressed images and not fail excessively for a subset of images. The Kappa coefficient is a measure of agreement. Usually, a Kappa coefficient superior to 0.4 is a good value; so the proposed blocking metric obtains a good agreement between subjective and predicted scores. Furthermore, the computational cost of the proposed blocking measure is less expensive: the proposed metric only uses the spatial information contained in the processed image, while the Wang’s blocking measure is based on the power spectrum of the 1-D absolute difference signal.

3.2. Conjoint experimental validation of ringing and blurring measures

In previous works [23], the proposed conjoint distortion measure of blurring and ringing effects did not take into account the weights introduced by the importance map in the computation of the different distortion measures. It included two parts: the first one was formulated as a linear combination of the first order terms of the blurring and ringing measures, plus the associated crossed term. The second part introduced an image feature (EM) defined by the percentage of strong edges contained in the compressed image. For images containing a few edges, the human visual perception of impairments is less sensitive. The use of an image feature allowed to the pooling model to take into account this human comportment. Experimental results demonstrated the efficiency of the proposed quality metric. In [20], we have demonstrated that the use of a computational attention model allows to increase significantly the correlation results. The Table 2 presents these results, performed using an image database [21] composed of 29 original high-resolution 24-bits/pixel RGB color images (typically 768x512) and their JPEG-2000 compressed versions with different compressed ratios.

<table>
<thead>
<tr>
<th>Model</th>
<th>Pearson Correlation</th>
<th>Spearman Correlation</th>
<th>Outlier Ratio</th>
<th>Kappa Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.89637</td>
<td>0.8716</td>
<td>6.20%</td>
<td>0.47957</td>
</tr>
<tr>
<td>2</td>
<td>0.94484</td>
<td>0.8956</td>
<td>4.08%</td>
<td>0.71183</td>
</tr>
</tbody>
</table>

Table 2. Comparison of correlation results between a conjoint measure of blurring and ringing effects without attention model (model 1) and the same one, but using the attention model (model 2).

3.3. Experimental validation of the proposed No Reference metric

In order to confirm the efficiency of the different proposed distortion measures, two new image databases are used: the first one is composed of 50 images compressed with JPEG coding and the second one, with JPEG-2000 coding.

It consists of 10 original high-resolution 24-bits/pixel RGB color images (typically 512x512) of well-known images of the image processing research community ("lena", "boats", "clown", "isabelle", "fruits", "barbara", "plane", "house", "pimento", "mandrill") and their five respective versions compressed with different compression ratios. The corresponding quality scores were taken from a 5-grade scale from 1, the lowest quality to 5, the highest one. For the JPEG image database, the Person Correlation of PSNR is equal to 0.5956 and in the
case of JPEG-2000 image database, the Person Correlation of PSNR is equal to 0.8075.

The table 3 presents the Pearson and Spearman correlations of each distortion measure, performed on the two different images databases. As envisaged, in the case of JPEG image database, the proposed blocking and blurring measures obtain the best correlations results and in the case of JPEG-2000 image database, the blurring and ringing measures are the best efficient, which is coherent with the human perception of different impairments respectively introduced by JPEG and JPEG-2000 coding.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking</td>
<td>0.8507</td>
<td>0.8836</td>
<td>0.2512</td>
<td>0.2519</td>
</tr>
<tr>
<td>Blurring</td>
<td>0.7207</td>
<td>0.7289</td>
<td>0.8234</td>
<td>0.8505</td>
</tr>
<tr>
<td>Ringing</td>
<td>0.3700</td>
<td>0.3870</td>
<td>0.8400</td>
<td>0.8089</td>
</tr>
</tbody>
</table>

Table 3. Correlation results of each proposed distortion measure, performed on the JPEG and JPEG-2000 image databases.

In our experiments, each database is randomly divided into two sets: 5 training images and 5 testing images, together with their compressed versions. The weights $a_i$ of the inter-distortion pooling model are estimated from training images using minimal mean squared error estimate between quality predictions and subjective scores. Then, the proposed trained quality metric is validated on the test database. The quality predictions resulting from this assessment are compared with human scores. In the case of JPEG-2000 image database, the computed weights $a_i$ associated to the blocking measure obtain lowest values. This is coherent, because the blocking effect is not classified as one of most important visual impairments introduced by JPEG-2000 coding.

The table 4 presents the correlation results of the proposed No Reference quality metric and the Wang’s metric [24], using the JPEG image database and the table 5 presents the correlation results of the proposed one and the Sheikh’s metric [14], using the JPEG-2000 image database.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Pearson Correlation</th>
<th>Spearman Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang’s metric</td>
<td>0.9392</td>
<td>0.9472</td>
</tr>
<tr>
<td>Proposed NR quality metric</td>
<td>0.9151</td>
<td>0.9102</td>
</tr>
</tbody>
</table>

Table 4. Comparison of correlation results between Wang’s metric and the proposed No Reference quality measure, computed using the JPEG image database.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Pearson Correlation</th>
<th>Spearman Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheikh’s metric</td>
<td>0.7136</td>
<td>0.7728</td>
</tr>
<tr>
<td>Proposed NR quality metric</td>
<td>0.8548</td>
<td>0.8502</td>
</tr>
</tbody>
</table>

Table 5. Comparison of correlation results between Sheikh’s metric and the proposed No Reference quality measure, computed using the JPEG-2000 image database.

4. CONCLUSION

In this paper, we have presented a new reference free quality metric for assessing the quality of JPEG and JPEG-2000 compressed images. The proposed approach is based on separate models specifically tuned to certain type of distortion (blocking, blurring, ringing), and combines their results according to the impact of each type of impairment on the image quality. The quality metric is validated on image databases assessed by human observers and is compared with two different metrics; each one specialized in the assessment of images compressed by JPEG and JPEG-2000 coding. For testing the prediction quality of the proposed metric, the performance tests approved by VQEG are satisfied, which demonstrates the efficiency of the proposed algorithm.

Currently, a region-based attention model is used in order to simulate the processing taking place in the HVS. The generated importance map allows to weight the ringing and blurring measures. However, the blocking measure does not take into account these weights: the computational attention model generates importance regions, while the blocking measure is computed using a block matching. Moreover, previous works have shown that the use of an attention model allows to increase significantly the efficiency of prediction quality. Our
future works aim to extend this approach, using a multi-resolution-based attention model, for weighting each one of proposed distortion measures.

5. REFERENCES