Novel Image Quality Metric Based on Similarity

Lina Jin¹, Nikolay Ponomarenko², Karen Egiazarian¹
¹Department of Signal Processing, Tampere University of Technology, Tampere, Finland
²Department of Signal Receivers, Transmitters and Signal Processing, National Aerospace University, Kharkov, Ukraine
lna.jin@tut.fi, karen.egiazarian@tut.fi
nikolay@ponomarenko.info

Abstract—In this paper, we present a novel approach to image quality metric taking into account degradation of contrast and brightness as well as block similarity. The metric is achieved by performing the following steps: 1) reducing contrast and brightness in distorted image, 2) using block-matching (BM) to group similar 2D image fragments into 3D data arrays in original image and preprocessed distorted image separately, 3) analyzing of these blocks in DCT domain. The DCT coefficients differences are calculated between pixel values with contrast sensitivity function (CSF) and reduced by contrast masking according to Human Visual System (HVS). We validate the performance of our algorithms with five most popular quality image databases: TID, LIVE, CSIQ, IVC and Cornell-A57. The analysis of the results shows that the proposed quality metric provides better correlation to Mean Observer Score (MOS) than most of recent popular state-of-the-art metrics, e.g. MSSIM, SSIM. The average Spearman Correlation of proposed metric reaches 0.894.

I. INTRODUCTION

Quality assessment (QA) plays an important role in the field of image processing applications. Subjective metrics evaluate the quality of the images by viewers and they are broadly used, e.g. MOS, DMOS [1]. However, such subjective assessments are quite expensive and time-consuming. Objective quality assessment (QA) aims to design an algorithm or metric to measure the quality of an image automatically and it should be acceptable and close to the human judgment. Peak signal-to-noise ratio (PSNR), and related to it, mean squared error (MSE) and root mean squared error (RMSE), widely used full-reference quality metrics, often show poor correlation with a perceived visual quality.

Recently, more and more quality metrics incorporate Human Visual System (HVS) properties (Contrast Sensitivity Function (CSF) and the luminance masking [2, 3]), making new measurements to be more consistent with a human visual perception. HVS perceive images by typically separating them into sub-bands or channels that are selective for spatial and temporal frequency as well as orientation. In [4], the Discrete Cosine Transform (DCT) has been utilized in contrast masking due to its suitability for certain applications and accuracy in modeling the cortical neurons. In the DCT domain, there are different approaches to model the contrast sensitivity masking in order to compute a visually optimal quantization matrix for a given image. Based on PSNR-HVS [4], PSNR-HVS-M [5] takes into account between-coefficient contrast masking of DCT basis the contrast sensitivity function. It operates with the values of DCT coefficients of 8x8 pixel block of an image. For each DCT coefficient of the block, the model allows to calculate its maximal distortion still not perceived, due to between-coefficient masking. Paper [6] proposed an efficient weighted MSE based metric which requires one reference image and two processed distorted images obtained by filtering the distorted image. The novel metric takes into account degradation of contrast and brightness as well as similar blocks based on models described in [4, 5, 6]. Comparing differences between grouped correlated similar blocks in DCT domain can achieve highly sparse representation of the quality of image.

The rest of the paper is organized as follows. In Section II the proposed model is described. The used set of test images is given in Section III. Section IV describes the experiment results and data analysis. Final conclusion is presented in Section V.

II. DESCRIPTION OF PROPOSED MODEL

A basic flow chart of the proposed model is presented in Figure 1. It mainly contains of four parts: 1. image preprocessing to process the distorted image by filtering, 2. apply block-match algorithm (BM) to group similar blocks in both reference image and distorted image and analyze the difference in DCT domain with CSF [4] and contrast masking [5], 3. Calculate modified MSE taking into account similar blocks.

A. Degradation of contrast and brightness

When images are captured by digital devices such as a camera, the quality will be influenced by mean shift (or intensity shift) and contrast change noise due to the light. For example, the pictures may look a little darker or brighter. This distortion information is present mainly in image homogenous regions. Compared to some other image features such as edges and textures, this difference may not affect much people to judge image quality according to human local contrast sensitivity properties, but will highly influence to some full reference metrics, such as PSNR, MSE.

To reduce this kind of errors, we process the distorted image $I_{\text{dis}}$ based on HVS local contrast sensitivity properties as following:

$$I'_{\text{dis}} = I_{\text{dis}} + \delta$$

$$I''_{\text{dis}} = \mu_{\text{dis}} + (I'_{\text{dis}} - \mu_{\text{dis}}) \frac{\sigma_{xy}}{\sigma_y^2}$$

where $I'_{\text{dis}}$, $I''_{\text{dis}}$ are the degradation brightness and contrast image respectively, $\delta$ is the mean difference between original
image $I_{org}$ and distorted image $I_{dis}$, $\mu_{dis}$ is the mean of $I_{dis}$, $\sigma_{xy}$ is the covariance of between $I_{org}$ and $I_{dis}$, $\sigma^2_x$ is the variance of $I_{dis}$.

**B. Block Similarity**

To find similar 2D fragments (blocks) in original image, the block-matching (BM) is applied as described in [7]. Block-matching algorithm is a particular matching approach that has been extensively used for motion estimation in video compression standards, such as MPEG 1, MPEG 2, H.264. BM is also used to group similar blocks and has been applied in many other applications, such as image denoising and image segmentation. The purpose of a block-matching algorithm is to find matching blocks, which are then stacked together in a 3D array [7]. An illustrative example of grouping by block-matching for images is given in Figure 2, which shows a reference block $A_0$ and the ones matched $A_i$ $(i = 1,2,\ldots,N)$ as similar to it within a macro window.

Similarity between image blocks is typically computed as the inverse of some distance measure. The smaller distance implies the higher similarity. Here, we use weighted Euclidean distance $l^2$-norm to calculate the distance. Pearson Correlation is used to measure the degree of association between reference block and searching block. A positive value for the correlation implies a positive association and a negative or inverse association. To find similar 2D fragments (blocks) in original image, the block-matching (BM) is applied as described in [7]. block-matching algorithm is a particular matching approach that has been extensively used for motion estimation in video compression standards, such as MPEG 1, MPEG 2, H.264. BM is also used to group similar blocks and has been applied in many other applications, such as image denoising and image segmentation. The purpose of a block-matching algorithm is to find matching blocks, which are then stacked together in a 3D array [7]. An illustrative example of grouping by block-matching for images is given in Figure 2, which shows a reference block $A_0$ and the ones matched $A_i$ $(i = 1,2,\ldots,N)$ as similar to it within a macro window.

Similarity between image blocks is typically computed as the inverse of some distance measure. The smaller distance implies the higher similarity. Here, we use weighted Euclidean distance $l^2$-norm to calculate the distance. Pearson Correlation is used to measure the degree of association between reference block and searching block. A positive value for the correlation implies a positive association and a negative or inverse association.

The searched blocks $A_i$ $(i = 1,2,\ldots)$ are ordered in an ascending order sequence according to the dissimilarities (i.e. distance) from the reference block $A_0$. The first $N_{sim}$ blocks are grouped, i.e. $A = \{A_i | i = 0,1,2,\ldots,N_{sim}\}$ from reference image $I_{org}$. In preprocessed distorted images $I_{dis}$ and $I_{dis}$, which have been degraded of brightness and contrast, $B = \{B_i | i = 0,1,2,\ldots,N_{sim}\}$ $C = \{C_i | i = 0,1,2,\ldots,N_{sim}\}$ can be directly grouped which have the same position of $A$ in the original image. For which have positive correlation to the reference block $A_0$ are the grouped $N_{sim}$ blocks $A_i$ $(i = 1,2,\ldots,N_{sim})$, only the ones considered mutually similar. Each block size is defined as 8x8 and it can be overlapped or not, either to reference block $A_0$ or searching blocks $A_i$ $(i = 1,2,\ldots)$. Here, we choose two similar blocks $(N_{sim} = 2)$ without an overlap to calculate efficiently. Then, DCT transform is applied to each grouped block of $A,B$ and $C$.

**C. Calculation of Modified MSE**

The modified version of MSE is calculated by the following formula:

$$MSE_H = \begin{cases} 
(MSE_c + K(MSE_b - MSE_c)), & \text{if } MSE_b > MSE_c \\
MSE_b, & \text{otherwise}
\end{cases} \quad (3)$$

$$MSE = MSE_H + 0.04 \times \delta^2 \quad (4)$$

where $MSE_b$ is the MSE of the brightness distortion $I_{dis}$, $MSE_c$ is the MSE of the contrast distorted image $I_{dis}$, parameter $K$ is set as

$$K = \begin{cases} 
0.002, & \text{if } \sigma_{xy} < 1 \\
0.25, & \text{otherwise}
\end{cases} \quad (5)$$

Both $MSE_b$ and $MSE_c$ are calculated taking into account similar blocks and HVS according to [5]:

$$MSE_b = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (e_{w}(A_0 - B_0) + \sum_{i=1}^{N_{sim}} \tau_i e_{w}(A_i - B_i))}{p} \quad (6)$$

$$MSE_c = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (e_{w}(A_0 - C_0) + \sum_{i=1}^{N_{sim}} \tau_i e_{w}(A_i - C_i))}{p} \quad (7)$$

$$\tau_i = \begin{cases} 
|\rho(A_0, A_i)|^a, & \text{if } \rho(A_0, A_i) > 0 \\
0, & \text{otherwise}
\end{cases} \quad (8)$$

$$P = (M - 7)(N - 7)(1 + \sum_{i=1}^{N_{sim}} \tau_i)64 \quad (9)$$

where $M, N$ denote image size, $e_{w}$ is given in [5], $A_0, B_0, C_0$ are the blocks which locate at index $(m,n)$ in $I_{org}$, $I_{dis}$, and $I_{dis}$ respectively, $N_{sim}$ is the number of similar blocks, $p$ is used to calculate the Pearson Correlation between similar block $A_i$ and reference block $A_0$, if $\rho(A_0, A_i) > 0$ we set the parameter $\tau_i$ as $\epsilon |\rho(A_0, A_i)|^a$. Here we set $\epsilon = 10, a = 0.05$ as default values obtained experimentally. Alternatively if $\rho(A_0, A_i) \leq 0$, then we set $\tau_i = 0$, i.e. the block is not considered.

Finally, modified PSNR is calculated using $MSE_H$:

$$sPHVS = 10 \log_{10}(255^2 / MSE_H) \quad (10)$$
III. TEST IMAGE SET

In our experiments we have used five most popular image databases: TID [8], LIVE [10], CSIQ [11], IVC [12] and Cornell-A57 [13].

TID [8] image database is the largest database of distorted images intended for verification of full-reference quality metrics. It contains 25 reference images and 68 distorted versions of each original image, in total 1700 distorted images. There are 17 types of different distortions including distortions most relevant for digital image processing applications. These 17 types of distortion are classified into 11 subsets, e.g. Noise, Noise2, Noise3, Safe, Hard, Simple, JPEG, Exotic, Exotic2, Exotic3 and Actual. Each subset includes one or more different types of distortions to simulate the situations which could occur in the digital image processing applications. For example, ‘JPEG’ includes the distortions due to compression using JPEG and JPEG2000 compression standards. For the detailed specifications of other subsets, please refer to [9]. For assessing the human perceived quality, 800 observers from three countries have carried out about 256000 individual human quality judgments (more than 300 judgments for each distorted image). The overall ratings of TID image database are presented as mean opinion score (MOS) [8].

The LIVE [10] image quality database is a publicly available test set of images, which contains 29 reference images and 779 distorted images. It is impaired with different artifacts and annotated with differential mean opinion scores (DMOS) values, which can be downloaded from [10]. The distortion types include JPEG2000 compression, JPEG compression, Gaussian noise contamination, Gaussian blur, and JPEG2000 compressed images undergoing fast fading channel distortions.

The CSIQ [11] database contains of 30 original images and 866 distorted versions of the originals. It mainly contains six types of distortion: JPEG, JPEG2000, global contrast decrements, additive pink Gaussian noise, additive white Gaussian noise, and Gaussian blurring. To each distortion type, there are four to five different levels of distortion. The subjective ratings were collected using a single stimulus absolute scaling. There are about 5000 subjective ratings from 25 different observers. The overall ratings of CSIQ are presented in the form of DMOS.

IVC [12] contains 10 original images and 235 distorted images which were generated from 4 different processing: JPEG, JPEG2000, LAR coding, and Blurring. Subjective evaluations were made at viewing distance of 6 times the screen height using a DSIS (Double Stimulus Impairment Scale) method with 5 categories and 15 observers. Distortions for each processing and each image have been optimized in order to uniformly cover the subjective scale.

Cornell-A57 [13] database is a small image quality database. It contains three original images, impaired by six types of distortions with varying parameters, in total 57 images. The distortions are: Flat allocation (FLT), Baseline JPEG compression (JPG), baseline JPEG2000 compression (JP2), JPEG2000 and Dynamic Contrast-Based Quantization (DCQ), Additive White Gaussian noise; blurring by using a Gaussian filter.


d| Spearman Correlation | Kendall Correlation |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sPHVS M</td>
<td>sPHVS M</td>
</tr>
<tr>
<td>Noise</td>
<td>0.913</td>
</tr>
<tr>
<td>Noise2</td>
<td>0.937</td>
</tr>
<tr>
<td>Noise3</td>
<td>0.919</td>
</tr>
<tr>
<td>Safe</td>
<td>0.925</td>
</tr>
<tr>
<td>Hard</td>
<td>0.817</td>
</tr>
<tr>
<td>Simple</td>
<td>0.935</td>
</tr>
<tr>
<td>JPEG</td>
<td>0.962</td>
</tr>
<tr>
<td>Exotic</td>
<td>0.800</td>
</tr>
<tr>
<td>Exotic2</td>
<td>0.831</td>
</tr>
<tr>
<td>Exotic3</td>
<td>0.650</td>
</tr>
<tr>
<td>Actual</td>
<td>0.919</td>
</tr>
<tr>
<td>Full</td>
<td>0.862</td>
</tr>
</tbody>
</table>

The result of subjective experiments consists in getting the ordered set of distorted test images. For analysis of adequacy of the considered metrics we have used Spearman rank-order correlations that can be exploited for determination of correlation between sorted data.

To compare performance of different objective metrics, we have used IVQEST tool. IVQEST (Image and Video QUality Evaluation SofTware) is proposed in [14] by IVU lab (The Image, Video, and Usability Laboratory) for evaluating image and video quality. It is a MATLAB based framework and provides a rich set of tools for objective metrics evaluation, test content generation, correlation analysis, subjective testing and subjective score analysis.

Several state-of-the-art image quality metrics are selected to compare the performance: MSSIM [15], SSIM [16], UQI [17], NQM [18], VIF [19], PSNR-HVS, PSNR-HVS-M, wPHVS [6], wPHVSM [6]. All algorithms use grayscale versions or luminance components of reference image and distorted images. Both original and distorted RGB images are transformed into grayscale firstly according to $I = 0.299R+0.589G+0.114B$, where R, G and B are the Red, Green and Blue components in RGB color space. The default implementation is provided in [14].

Table I shows the Spearman correlation and Kendall correlation for the presented metric and MSSIM to TID database. Figures 3 presents the results of fitting, ‘sPHVSM’ and ‘sPHVS’ are the proposed metric used with and without masking effect [5], respectively. As mentioned above, distorted images in TID are grouped in 11 subsets. Block similarity in proposed model groups the similar blocks using BM and reduces the error of ‘Non eccentricity pattern noise’ and ‘Local block-wise distortions of different intensity’ [9], which may be produced in image compression, watermarking, image acquisition and inpainting. The above two distortion types are contained in all ‘Exotic’ series subsets. The proposed metrics (with and without masking effect) also effectively improve the distorted images of ‘Mean Shift’ and ‘Contrast change’ by degradation of contrast and brightness. Compared with the
Spearman correlations have been obviously improved for ‘Exotic’, ‘Exotic2’ and ‘Exotic3’. The final Spearman and Kendall correlations of full database are higher than those for MSSIM.

Table II shows the Spearman Correlations of compared image quality metrics on five image quality databases: TID, LIVE, CSIQ, IVC and A57. sPHVSM and sPHVS perform more stable on different databases than PSNR-HVS-M and PSNR-HVS. The average Spearman correlation of sPHVSM is 0.894 and 0.881 to sPHVS. The metrics sPHVSM and sPHVS which take into accounting mean shift, contrast changing and block similarity provide better correspondence to human perception than ones (wPHVSM and wPHVS) which allow for only mean shift and contrast changing.

V. CONCLUSIONS

In this paper, we present a novel approach to image quality metric taking into account degradation of contrast and brightness as well as block similarity based on PSNR-HVS and PSNR-HVS-M. From the experiment results (Section IV), it can be seen that the presented model has significantly improved the correlation to MOS. It effectively improve the performances to distorted images such as ‘Non eccentricity pattern noise’, ‘Local block-wise distortions of different intensity’, ‘Mean Shift’ and ‘Contrast Change’ compared to PSNR-HVS and PSNR-HVS-M. Proposed metric outperforms most of current state-of-the-art image quality metrics.

REFERENCES


