A NOVEL SVD-BASED IMAGE QUALITY ASSESSMENT METRIC

Shuigen Wang\textsuperscript{1}, Chenwei Deng\textsuperscript{1}, Weisi Lin\textsuperscript{2}, Baojun Zhao\textsuperscript{1}, Jie Chen\textsuperscript{3}

\textsuperscript{1} School of Information and Electronics, Beijing Institute of Technology, China
\textsuperscript{2} School of Computer Engineering, Nanyang Technological University, Singapore
\textsuperscript{3} School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

ABSTRACT

Image distortion can be categorized into two aspects: content-dependent degradation and content-independent one. An existing full-reference image quality assessment (IQA) metric cannot deal with these two different impacts well. Singular value decomposition (SVD) as a useful mathematical tool has been used in various image processing applications. In this paper, SVD is employed to separate the structural (content-dependent) and the content-independent components. For each portion, we design a specific assessment model to tailor for its corresponding distortion properties. The proposed models are then fused to obtain the final quality score. Experimental results with the TID database demonstrate that the proposed metric achieves better performance in comparison with the relevant state-of-the-art quality metrics.

Index Terms— Image Quality Assessment, Human Visual System, Singular Value Decomposition

1. INTRODUCTION

Image quality assessment (IQA) has played an important role in numerous applications, such as image acquisition, watermarking, transmission, compression, restoration, enhancement, etc. The most reliable way for IQA is the subjective evaluation; however, it is too cumbersome and expensive to be used in computational and automated systems. Hence, an objective image quality metric consistent with the subjective perception is in demand.

The simplest and most widely used quality assessment metrics are the mean-square-error (MSE) and its corresponding peak signal-to-noise ratio (PSNR), which are simple to compute and have clear physical meanings. However, since these two metrics pay no attention to image structure, they are widely criticized for not correlating well with the human perception, though they are good for low content-dependent distortions (e.g., random additive noise).

Since the human visual system (HVS) is sensitive to structural information of the input scene, a number of HVS-based IQA metrics have been proposed during recent few years, such as the structural similarity (SSIM) \cite{1}, the multi-scale SSIM index (MSSIM) \cite{2}, the visual information fidelity (VIF) \cite{3}, the visual signal-to-noise ratio (VSNR) \cite{4}, the most apparent distortion (MAD) \cite{5}, the mean geometric structural distortion (MGSD) \cite{6}, the feature similarity (FSIM) \cite{7}, and the recently proposed visual gradient similarity (VGS) \cite{8}.

However, different types of distortion have different properties, e.g., the additive white Gaussian noise (awgn) is a high frequency impairment to the images; the blur degrades image details while maintaining the major structures; and the JPEG and JPEG2000 compression distortion induce some detailed information loss and additional structural noise (e.g., blockiness). There is no single existing metric suitable for all types of distortion. Some metrics mainly consider the structural characteristics (e.g., gradient/edge) in their formulation, such as the SSIM, FSIM, MGSD and VGS, while the PSNR/MSE have good correlation with the subjective perception on the content-independent distortion.

Unlike the previous metrics mentioned above, we aim at designing a comprehensive metric to adapt to as many types of distortion as possible. The basic idea is that, different distortion types are firstly analyzed and separated; for more consistent assessment in line with the HVS perception, different measurement methods are then utilized to exploit the properties of corresponding distortion; and these measures are pooled together to deduce the overall quality score.

2. MOTIVATION

As reported in the TID database \cite{9}, there are 17 types of distortion that have been widely investigated, e.g., awgn, blur, jpeg compression, and transmission error, etc. From the discussion above, there are two major types of distortion: content-dependent degradation and content-independent degradation. Generally, image impairments are the combination of the two categories. It is more reasonable to evaluate the two categories of distortion separately, because the responses of the HVS toward them are not the same.

The SVD is an effective tool for feature extraction and image decomposition. In \cite{10}, Manish et al. used SVD to extract image features and applied machine learning for feature pooling, without separating the two aforementioned distortion aspects. Liu et al. \cite{11} advocated that for images distorted by awgn, the image signal contributes much to the early part of
3. PROPOSED ALGORITHM

In this section, we are to make detailed explanation for the proposed IQA model. We extract the information of structure and that of the less content-dependent component from the reference (and test) images via SVD, and the resultant two portions of information are separately evaluated.

3.1. Image Decomposition

The SVD of an $m \times n$ image matrix $X$ (assume the rank of $X$ is $q$) yields three matrices: an orthogonal matrix $U$, a diagonal matrix $\Delta$, and the transpose of another orthogonal matrix $V$. It can be represented as

$$X = U \times \Delta \times V^T,$$

where $U^TU = I_{mm}$, $V^TV = I_{nn}$ ($I_{mm}$ and $I_{nn}$ are the m-square and n-square identity matrices), and

$$U = [u_1 \ u_2 \ \cdots \ u_m],$$

$$V = [v_1 \ v_2 \ \cdots \ v_n],$$

$$\Delta = diag(\delta_1 \ \delta_2 \ \cdots \ \delta_q),$$

where $u_i$ and $v_j$ are column vectors of $U$ and $V$, and $\delta_k$ is a singular value ($i = 1, 2, \cdots, m, j = 1, 2, \cdots, n, k = 1, 2, \cdots, q, q = rank(X)$). Besides, the singular values appear in descending order, i.e., $\sigma_1 > \sigma_2 > \cdots > \sigma_q$.

Each term $X_k = u_k \delta_k v_k^T$ ($k = 1, \cdots, q$) can be regarded as a basis image, and it specifies a layer of the image geometry. The content-dependent component is dominated by the first a few basis images, while the less content-dependent information is determined by the remaining ones.

where $q$ is the rank of input image (following the notation in (2)), $p \in (0 \ q)$, the selection of $p$ is based on the image size and distortion type; $R_r$ (or $R_d$) and $I_r$ (or $I_d$) are the content-dependent part and low content-dependent (or content-independent) part of the reference (or distorted) image, respectively.

$$R_r = \sum_{k=1}^{p} u_k \delta_k v_k^T,$$

$$I_r = \sum_{k=p+1}^{q} u_k \delta_k v_k^T,$$

$$R_d = \sum_{k=1}^{p} u_k \delta_k v_k^T,$$

$$I_d = \sum_{k=p+1}^{q} u_k \delta_k v_k^T,$$

3.2. Degradation Evaluation

For the content-dependent part, we follow the concept of the works in [1, 12] and adopt gradient and contrast similarity to compute the quality degradation. The gradient similarity between the content-dependent information of the reference image ($R_r$) and the distorted image ($R_d$) is defined as

$$g(x, y) = \frac{2G_{R_r}(x)G_{R_d}(y) + C_1}{G_{R_r}(x)^2 + G_{R_d}(y)^2 + C_1}.$$
where $G_{R_c}(x)$ and $G_{R_c}(y)$ are the gradient values of the central pixels of image blocks $x$ and $y$, $C_1$ is a small constant to avoid the denominator being zero and is set as $C_1=(0.03 \times L)^2$, here $L$ is the gray level of the image.

The gradient value $G_{R_c}(x)$ (similar for $G_{R_c}(y)$) is computed as the maximum response along the four directions,

$$G_{R_c}(x) = \max_{k=1,...,4} |\varphi M_k \ast R_c|,$$
where $M_k$ ($k = 1,2,3,4$) are four filter operators as shown in Fig. 3, $\varphi$ is a constant and empirically can be set as $1/16$ [12], symbol $\ast$ denotes the convolution operation.

Apart from the discussion above, there exists some structural information having no apparent edge but strong contrast difference. To measure the degradation on image contrast information, we adopt contrast similarity [1] as follows

$$c(x, y) = \frac{2\sigma_{R_c}(x)\sigma_{R_c}(y) + C_2}{\sigma_{R_c}(x)^2 + \sigma_{R_c}(y)^2 + C_2},$$
where $\sigma_{R_c}(x)$ and $\sigma_{R_c}(y)$ are the standard deviations of the image blocks $x$ and $y$, respectively, and $C_2=C_1/2$ [1].

Combining the gradient and contrast similarities, the evaluation of content-dependent impact can be formulated as

$$s(x, y) = g(x, y) \cdot c(x, y),$$
(8)

For content-independent component, the MSE-like metrics correlate well with the subjective ratings [13]. Here we employ normalized PSNR to evaluate the quality of this part,

$$N(I_r, I_d) = \frac{1}{C_3}10\log_{10} \left( \frac{255^2}{\text{MSE}(I_r, I_d)} \right),$$
(9)
where $I_r$ and $I_d$ are the content-independent parts of the reference and distorted images, respectively, and $\text{MSE}(I_r, I_d)$ is the mean squared error between $I_r$ and $I_d$ (to avoid the denominator zero, we set the minimal value of MSE calculation as 1); and we set $C_3=10\log_{10}255^2$ to normalize the PSNR value into the range of $[0, 1]$.

### 3.3. The Overall Quality Evaluation

Degradations on content-dependent and content-independent parts co-determine the image quality. Therefore, we combine the evaluation results of the two impacts using a non-linear expression to acquire the perceptual quality score.

$$Q = S(R_r, R_d)\alpha \cdot N(I_r, I_d)\beta,$$
(10)
where $S(R_r, R_d)$ is the pooling value of the content-dependent impact (the mean value of all $s(x, y)$ in equation (8)); the parameters $\alpha$ and $\beta$ are adaptive to the relative importance of the two parts, as to be determined next.

As aforementioned, the SVD of the image yields a singular value matrix $\Delta$ in the descending order; each singular value corresponds to a basis image. In fact, the singular value represents the importance of its relevant basis image. The bigger the value is, the more important the basis image is. The weights for the evaluation for the two parts are determined by the sum of their corresponding singular values,

$$\alpha = \frac{\sum_{i=1}^p \delta_i}{\sum_{i=1}^q \delta_i},$$
(11)

where the numerator is the sum of singular values of content-dependent portion, while the denominator denotes the sum of singular values of the original image. Moreover, since the singular values are non-negative, $\alpha \in [0, 1]$. Therefore, we set $\beta=1-\alpha$.

### 4. EXPERIMENTAL RESULTS

In this section, the effectiveness of proposed metric is verified in terms of its ability to predict quality in a manner that agrees with subjective perception. We analyze the performance of the proposed metric in the TID database [9], which is the most comprehensive publicly available one so far with 1,700 512 × 384 images. The database contains 25 reference GRB images, and each reference image corresponds to 17 distortion types with 4 distortion levels. Moreover, the mean opinion score (MOS) of each distorted image is provided.

To evaluate the performance of the IQA metrics in a common space, a five-parameter logistic mapping between the objective outputs $S_o$ and subjective scores is employed following the Video Quality Experts Group Phase-I/II tests,

$$S_m = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(S_o - \beta_3))} \right) + \beta_4 S_o + \beta_5,$$
(12)
where $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ are the parameters to be fitted by minimizing the sum of squared differences between the mapped values $S_m$ and the ground truth values (MOS) based on the Spearman rank-order correlation coefficient (SRC-C), Pearson linear correlation coefficient (PLCC) and root mean squared error (RMSE). The SRCC can measure prediction monotonicity and the other two are used to evaluate prediction accuracy. We employ the SRCC criteria in our experiments, since the other two lead to similar conclusions.

In our experiments, the parameter $p$ (in equation (4)) is set as 64. We compare the proposed metric with VGS [8], FSIM [7], MAD [5], VIF [3], VSNR [4], MSSIM [2], SSIM [1], PSNR, and the existing SVD-based metric (SVDR) [10]. The simulation results are listed in Table 1, and the two best IQA metrics have been highlighted in boldface.
Table 1. COMPARISON OF SRCC VALUES OF IQA METRICS FOR EACH DISTORTION TYPE

<table>
<thead>
<tr>
<th>Distortion</th>
<th>Proposed</th>
<th>VGS</th>
<th>SVDR</th>
<th>FSIM</th>
<th>MAD</th>
<th>VIF</th>
<th>VSNR</th>
<th>MSSIM</th>
<th>SSIM</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>awgn</td>
<td>0.8860</td>
<td>0.8850</td>
<td>0.7600</td>
<td>0.8566</td>
<td>0.8388</td>
<td>0.8799</td>
<td>0.7728</td>
<td>0.8094</td>
<td>0.8107</td>
<td>0.9114</td>
</tr>
<tr>
<td>awgn-color</td>
<td>0.8827</td>
<td>0.8940</td>
<td>0.7203</td>
<td>0.8527</td>
<td>0.8258</td>
<td>0.8785</td>
<td>0.7793</td>
<td>0.8064</td>
<td>0.8029</td>
<td>0.9068</td>
</tr>
<tr>
<td>spatial cor-noise</td>
<td>0.8734</td>
<td>0.8960</td>
<td>0.7875</td>
<td>0.8483</td>
<td>0.8678</td>
<td>0.8703</td>
<td>0.7665</td>
<td>0.8195</td>
<td>0.8144</td>
<td>0.9229</td>
</tr>
<tr>
<td>masked noise</td>
<td>0.8596</td>
<td>0.7860</td>
<td>0.6363</td>
<td>0.8021</td>
<td>0.7336</td>
<td>0.8698</td>
<td>0.7295</td>
<td>0.8155</td>
<td>0.7795</td>
<td>0.8487</td>
</tr>
<tr>
<td>high-fre-noise</td>
<td>0.9121</td>
<td>0.9360</td>
<td>0.8638</td>
<td>0.9093</td>
<td>0.8864</td>
<td>0.9075</td>
<td>0.8811</td>
<td>0.8685</td>
<td>0.8729</td>
<td>0.9323</td>
</tr>
<tr>
<td>impulse noise</td>
<td>0.7999</td>
<td>0.7260</td>
<td>0.6630</td>
<td>0.7452</td>
<td>0.6499</td>
<td>0.8331</td>
<td>0.6471</td>
<td>0.6868</td>
<td>0.6732</td>
<td>0.9177</td>
</tr>
<tr>
<td>quantization noise</td>
<td>0.8912</td>
<td>0.8700</td>
<td>0.8130</td>
<td>0.8564</td>
<td>0.8160</td>
<td>0.7956</td>
<td>0.8270</td>
<td>0.8537</td>
<td>0.8531</td>
<td>0.8699</td>
</tr>
<tr>
<td>blur</td>
<td>0.9706</td>
<td>0.9030</td>
<td>0.8120</td>
<td>0.9472</td>
<td>0.9197</td>
<td>0.9546</td>
<td>0.9396</td>
<td>0.9320</td>
<td>0.9607</td>
<td>0.9544</td>
</tr>
<tr>
<td>denoising</td>
<td>0.9720</td>
<td>0.9650</td>
<td>0.8893</td>
<td>0.9603</td>
<td>0.9434</td>
<td>0.9189</td>
<td>0.9286</td>
<td>0.9571</td>
<td>0.9530</td>
<td>0.9381</td>
</tr>
<tr>
<td>jpg-comp</td>
<td>0.9712</td>
<td>0.9700</td>
<td>0.8855</td>
<td>0.9279</td>
<td>0.9275</td>
<td>0.9170</td>
<td>0.9174</td>
<td>0.9348</td>
<td>0.9252</td>
<td>0.9011</td>
</tr>
<tr>
<td>jpg2k-comp</td>
<td>0.9800</td>
<td>0.9760</td>
<td>0.9027</td>
<td>0.9773</td>
<td>0.9707</td>
<td>0.9713</td>
<td>0.9515</td>
<td>0.9736</td>
<td>0.9625</td>
<td>0.8300</td>
</tr>
<tr>
<td>jpg-trans-error</td>
<td>0.8504</td>
<td>0.8290</td>
<td>0.8347</td>
<td>0.8708</td>
<td>0.8661</td>
<td>0.8582</td>
<td>0.8056</td>
<td>0.8736</td>
<td>0.8678</td>
<td>0.7665</td>
</tr>
<tr>
<td>jpg2k-trans-error</td>
<td>0.8902</td>
<td>0.7928</td>
<td>0.8653</td>
<td>0.8544</td>
<td>0.8394</td>
<td>0.8510</td>
<td>0.7909</td>
<td>0.8525</td>
<td>0.8577</td>
<td>0.7765</td>
</tr>
<tr>
<td>pattern-noise</td>
<td>0.6943</td>
<td>0.8060</td>
<td>0.6600</td>
<td>0.7491</td>
<td>0.8287</td>
<td>0.7608</td>
<td>0.5716</td>
<td>0.7336</td>
<td>0.7107</td>
<td>0.5931</td>
</tr>
<tr>
<td>block-distortion</td>
<td>0.8611</td>
<td>0.7790</td>
<td>0.8013</td>
<td>0.8492</td>
<td>0.7970</td>
<td>0.8320</td>
<td>0.8126</td>
<td>0.7617</td>
<td>0.8462</td>
<td>0.5852</td>
</tr>
<tr>
<td>mean shift</td>
<td>0.4969</td>
<td>0.5490</td>
<td>0.5152</td>
<td>0.6720</td>
<td>0.5161</td>
<td>0.5132</td>
<td>0.3715</td>
<td>0.7374</td>
<td>0.7231</td>
<td>0.6974</td>
</tr>
<tr>
<td>contrast</td>
<td>0.5508</td>
<td>0.8820</td>
<td>0.4360</td>
<td>0.6481</td>
<td>0.2723</td>
<td>0.8190</td>
<td>0.4239</td>
<td>0.6400</td>
<td>0.5246</td>
<td>0.6126</td>
</tr>
</tbody>
</table>

From Table 1, we can see that for the TID database, the proposed metric correlates well with the majority types of distortion, including awgn, quantization noise, blur noise, denoising, the jpg-comp, jpg2k-comp, jpg2k transmission error and block distortion, etc. PSNR is known to be good for additive noise situation, and as can be seen, the proposed metric is competitive to PSNR on the two additive noise (awgn and awgn-color), as well as the spatial correlated noise and high frequency noise. Overall, the proposed IQA metric performs the best among all state-of-the-art algorithms under comparison, and is highly consistent with the human perception.

5. CONCLUSION

In this paper, we proposed a novel SVD-based metric for image quality assessment. Image distortion can be categorized into content-dependent degradation and content-independent degradation, and their impacts are different to image quality evaluation. We investigated into the use of SVD to separate such two degradation types. Structure/contrast similarities and PSNR metric are then employed to evaluate the degradations on the two parts respectively. Finally, we combine the two evaluation results to obtain the overall quality score. Experimental results in the TID database have demonstrated that our metric outperforms the relevant existing algorithms and is highly consistent with the subjective ratings.

6. REFERENCES