Abstract In this paper we investigate the comparison of image quality assessment (IQA) methods for color images. An IQA evaluates a pair of images and returns a single number as a prediction of the perceived image difference. We introduce the Redundant Information Content Evaluation (RICE) that detects redundancies between distinct IQAs. Such comparisons help to develop or improve IQAs with respect to the information that they extract from the images.

RICE is analyzed using a set of common color IQAs applied on a large image difference database. Its functionality and performance is tested and an application for color image difference analysis is proposed.

1 Introduction

The assessment of image quality is an important task in many image processing fields, such as in image compression or reproduction. Since visual experiments with
human subjects are exhausting and expensive, an automatic computational evaluation of image quality is highly desired in imaging science. Thus, a computed image quality assessment (IQA) that accurately predicts human judgments is an active research area. The most common way of quality evaluation is the comparison of a distorted image with a reference image. An IQA transforms these images into a single number that predicts the perceived difference between the image pair. Some IQAs sum up the deviations from the reference image. A famous representative is the mean squared error (MSE). Also, signal-to-noise ratios (SNR) can be extracted from the image pair, e.g., peak SNR (PSNR). Other IQAs extract and compare information content from the images, such as structure, gradients, lightness, etc. The structural similarity (SSIM) index [1] makes use of this approach quite successfully. Using weighting factors or multi-scale methods as extensions of existing IQAs (e.g., weighted SNR (WSNR) [2] or multi-scale SSIM (MSSIM) [3], respectively) can even improve the prediction performance.

There exists a large variety of IQAs and still more are being developed. As long as the human visual system (HVS) is not fully understood these IQAs do not correlate perfectly with human perception. However, they rely on more or less accurate assumptions on image difference features that influence human judgments.

In this work we compare various IQAs with respect to the information content they extract. For each IQA we compute a vector of predictions for a large set of image pairs. These vectors intrinsically depend on the extracted image difference information. IQAs that rely on similar information are likely to show a higher correlation between their predictions than IQAs that are based on mutually different information. In this paper we investigate different approaches to detect such redundant information based on the IQA vectors. We call this methodology Redundant Information Content Evaluation (RICE).

We hope to provide a useful tool to evaluate the magnitude of information increment utilized by newly developed IQAs compared to existing approaches. This could be particularly helpful for selecting and combining different IQAs to improve the overall prediction performance. For instance, IQAs that utilize color attributes (e.g., lightness (L), chroma (C) and hue (h)) or different scales need an investigation of their individual parts. Furthermore, an IQA that utilizes information similar to that extracted by human observers to judge the difference of an image pair is well correlated to perception. Such judgments are obtained by visual experiments and their results are typically described by so called Mean Opinion Scores (MOS).
2 Methodology of Redundant Information Content Evaluation

A large number of predictions is necessary to apply RICE to the IQAs that we want to evaluate. Therefore, we take a large color image database with \( N \) distorted images and compute all IQA predictions for every image. Thus, we create \( N \)-dimensional column vectors for the investigated IQAs and denote them \( IQA_1, IQA_2, \ldots \).

In this paper we introduce five RICE methods:

1) **Condition number method**: The condition number (CN) method was proposed in [4] for selecting the most significant samples from a large number of samples. The condition number \( \text{Cond}(X) \) of a matrix \( X \) is the ratio of its largest to its smallest singular value. For the evaluation we create the matrix \( X \) from \( n \) different IQA vectors:

\[
X = [IQA_1, IQA_2, \ldots, IQA_n], \quad n \in \{2, 3, \ldots\}. \tag{1}
\]

Every row of \( X \) denotes a point in the \( n \)-dimensional space corresponding to a distorted image. We scale by \( n \) to account for the number of included IQAs:

\[
\text{CN} = \frac{\text{Cond}(X)}{n} \tag{2}
\]

A high condition number of \( X \) means that the points are not uniformly distributed within the \( n \)-dimensional space but predominantly within a lower dimensional subspace. Hence, the \( n \) IQAs contain some redundant information.

2) **Determinant of covariance method**: The covariance matrix \( \text{Cov}(X) \) specifies the variation between the involved IQAs. The determinant \( \text{Det}(X) \) of a matrix \( X \) is the product of the eigenvalues. To avoid very small values the determinant of covariance (DC) method is defined as

\[
\text{DC} = 10 \times (\text{Det}(\text{Cov}(X)))^{1/n}. \tag{3}
\]

A large DC value implies a high variation of the involved IQAs and therefore indicates low redundant information content.

3) **Partial correlation method**: The partial correlation \( r_{AB,C} \) is the correlation between two correlation vectors \( A \) and \( B \) with respect to a set of control vectors \( C \). These two vectors are projected onto the vector subspace that is orthogonal to the control vectors, i.e., the information of the control vectors is removed. In the special case that the control vectors only consist of a single
variable $C$ the partial correlation can be calculated with the Pearson correlation $r_{AB}$ [5]:

$$r_{AB,C} = \frac{r_{AB} - r_{AC} \cdot r_{BC}}{\sqrt{1 - r_{AC}^2} \cdot \sqrt{1 - r_{BC}^2}}$$

(4)

For the partial correlation (PC) method we set $A = IQA_1$, $B = IQA_2$ and $C = [IQA_3, IQA_4, ..., IQA_n]$. In the case of a pairwise IQA comparison, the Pearson correlation is used instead. A high partial correlation between the two correlation vectors $A$ and $B$ indicates a lot of redundant information. In MATLAB, partial correlation is implemented as the function “partialcorr”.

4) **Intraclass correlation method:** The intraclass correlation (ICC) is a modified Pearson correlation where the mean of each single IQA is replaced by the mean of all IQAs. In its original form [6] the ICC for two pairs $(A_{n,1}, A_{n,2})$ of $N$ variables with mean $\bar{A} = \frac{1}{2N} \sum_{n=1}^{N} (A_{n,1} + A_{n,2})$ is

$$ICC = \frac{2 \cdot \sum_{n=1}^{N} (A_{n,1} - \bar{A}) \cdot (A_{n,1} - \bar{A})}{\sum_{n=1}^{N} (A_{n,1} - \bar{A})^2 + \sum_{n=1}^{N} (A_{n,2} - \bar{A})^2}.$$  

(5)

For RICE, a low ICC indicates a high variance of the IQA predictions and therefore low redundancy. Modern versions of ICC can be found, for instance, on the MATLAB Central file exchange.

5) **Principle angle method:** The principle angle (PA) specifies the minimal angle between two vector subspaces. For our approach we determine the principle angle between a single vector $A = IQA_1$ and a vector subspace spanned by $B = \{IQA_2, ..., IQA_n\}$. Let $\hat{A} = A/\|A\|$ and $U$ be a matrix with orthonormal column vectors that span the same vector subspace than $B$. The principle angle PA is then calculated by

$$PA = \cos^{-1}(\text{Svd}(\hat{A}^t \cdot U)),$$

(6)

where $\text{Svd}(\hat{A}^t \cdot U)$ is the singular value of the matrix $\hat{A}^t \cdot U$. The columns of $U$ can be calculated using the Gram-Schmidt orthonormalization of $B$ (assuming $B$ is linearly independent). The smaller the principle angle, the closer is $A$ to the vector subspace spanned by $B$ and the more redundant is the information content.
3 Image Quality Assessments and Color Image Database for the Evaluation

For the evaluation of redundant information content we use a set of common IQAs provided by the MeTriX MuX package [7]. Additionally, we utilize two extensions of PSNR taking into account the human contrast sensitivity functions, PSNR-HVS and PSNR-HVSM [8]. All IQAs are listed in table 1 with their internal number used in this work.

<table>
<thead>
<tr>
<th>#</th>
<th>IQA</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MSE</td>
<td>PSNR</td>
<td>SSIM</td>
<td>MSSIM</td>
<td>VSNR</td>
<td>VIF</td>
<td>VIFP</td>
<td>UQI</td>
</tr>
<tr>
<td>#</td>
<td></td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IFA</td>
<td>IFC</td>
<td>NQM</td>
<td>WSNR</td>
<td>SNR</td>
<td>PSNR-HVS</td>
<td>PSNR-HVSM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: List of all IQAs used in this work with their internal number (#).

We choose the publicly available Tampere Image Database 2008 (TID2008) [9] for our analysis. It contains a large number of images (1700 distorted images based on 25 reference images) with a variety of distortions and distortion levels. Many distortions that typically occur in image processing applications, e.g., JPEG compression artifacts, blurring, Gaussian noise, etc., are included. In a visual experiment with 838 subjects the perceived difference of a distorted image to its reference image was determined by calculating the MOS. As an example, a reference image and three corresponding distorted images of this database are shown in fig. 1.

Reference image  Gaussian noise  Gaussian blur  JPEG compression

MOS = 0.473  PSNR = 0.719  SSIM = 0.599  VIF = 0.706
MOS = 0.650  PSNR = 0.719  SSIM = 0.205  VIF = 0.837
MOS = 0.708  PSNR = 0.672  SSIM = 0.441  VIF = 0.875

Figure 1: Example of TID2008: Three distorted images (Gaussian noise, Gaussian blur and JPEG compression) originating from a reference image and their MOS, PSNR, SSIM and VIF values.
4 Investigation of the proposed RICE methods

The five RICE methods lead to different results. Therefore, their performance and functionality have to be analyzed.

The RICE approach can be applied to a single pair of IQAs as well as to a whole group of IQAs. A pairwise methodology is outlined in fig. 2 for the evaluated set of IQAs on the color image database mentioned above. In a first step we compute the fourteen IQAs for every image of the database. In the next step we compare the IQAs with each other using RICE. Since we apply RICE to every possible pair of IQAs the result can be summarized in a matrix. Note that RICE does not have to be symmetric.

![Figure 2: Methodology of the pairwise RICE approach](image)

As a first investigation the functionality has to be verified. For this purpose three model IQAs were computed for every image:

1. The peak signal-to-noise ratio (PSNR) as an exemplary IQA,
2. PSNR with low random noise added and
3. A random number in the range of $[0,1]$. Then, they are compared by the pairwise RICE method. The results illustrated in table 2 show the expected behavior of RICE for redundant information content (PSNR vs. PSNR with noise) and non-redundant content (PSNR vs. random number), respectively. These RICE values can roughly be seen as the lower and upper limits for every method but they differ for different model IQAs, numbers $n$ of involved IQAs and image databases.

<table>
<thead>
<tr>
<th>Method</th>
<th>CN</th>
<th>DC</th>
<th>PC</th>
<th>ICC</th>
<th>PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR vs. PSNR with noise</td>
<td>668</td>
<td>0.001</td>
<td>1.000</td>
<td>1.000</td>
<td>0.001</td>
</tr>
<tr>
<td>PSNR vs. random number</td>
<td>2</td>
<td>0.136</td>
<td>0.036</td>
<td>0.012</td>
<td>0.524</td>
</tr>
</tbody>
</table>

Table 2: Results of the functionality test for redundant (PSNR vs. PSNR with noise) and non-redundant (PSNR vs. random number) information content of the five RICE methods.

The next investigation is the analysis of the pairwise RICE for all five methods as described in fig. 2. The **most redundant information** can be found, as expected, between IQAs based on similar approaches. These are the following two pairs leading to the most redundant RICE values for all methods: {PSNR,SNR} and {PSNR-HVS,PSNR-HVSM}. Apart from the DC method also {VIF,VIFP} and {WSNR,PSNR-HVSM} are found to contain a lot of redundant information.

PSNR and MSE should extract the same information content since they depend on each other:

$$PSNR = 20 \cdot \log \frac{I_{max}}{\sqrt{MSE}},$$  

(7)

with $I_{max}$ being the maximal signal intensity. However, none of the methods is able to adequately detect this kind of relation. Only the PC and ICC methods detect this relation as highly redundant but not as totally redundant. Therefore, these two methods should be preferred for a pairwise analysis of redundant information content.

The results of the proposed RICE methods for choosing the IQA pair with the **lowest redundant information** content differ widely. However, the results are reasonable for all methods since they comprise IQA pairs which are based on different approaches, such as {UQI,WSNR}, {VIFP,PSNR-HVSM}, etc.
5 Conclusion

We introduced five different methods for redundant information content evaluation of image quality assessments called RICE. They were applied to fourteen different IQA vectors computed on a set of color images. A first investigation verified the functionality of all methods and provided approximate upper and lower limits. In a pairwise comparison of IQAs, RICE detects redundancies between IQAs based on similar approaches. The partial correlation method as well as the intraclass correlation method yield the best performance. For pairs of IQAs with low information redundancy the RICE methods show different but reasonable results. The pairwise RICE methodology is useful for the detection of redundancies and therefore provides a useful tool for developing new IQAs with larger information increment or to find highly non-redundant IQAs that might be blend to an improved single IQA.

6 Acknowledgments

We would like to thank Ingmar Lissner for his help in writing this paper. The research was conducted with financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft).

References


