Image Quality assessment
Perceptual Image Processing

Standard measure (MSE) does not agree with human visual perception

PERCEPTUAL IMAGE PROCESSING

Define Perceptual IQA Measures
Optimize IP Systems & Algorithms “Perceptually”

Why?

Application Scope: essentially all IP applications
image/video compression, restoration, enhancement, watermarking, displaying, printing ...
Image Quality Assessment

• Goal
  — Automatically predict perceived image quality

• Classification
  — Full-reference (FR); No-reference (NR); Reduced-reference (RR)

• Widely Used Methods
  — FR: MSE and PSNR \( PSNR = 10 \log_{10} \frac{L^2}{MSE} \)
  — NR & RR: wide open research topic

• IQA is Difficult
VQEG (1)

• VQEG (video quality experts group)

1. Goal: recommend video quality assessment standards
   (TV, telecommunication, multimedia industries)

2. Hundreds of experts
   (Intel, Philips, Sarnoff, Tektronix, AT&T, NHK, NASA,
    Mitsubishi, NTIA, NIST, Nortel ……)

• Testing methodology

1. Provide test video sequences
2. Subjective evaluation
3. Objective evaluation by VQEG proponents
4. Compare subjective/objective results, find winner
VQEG (2)

• Current Status

1. Phase I test (2000):
   ▪ Diverse types of distortions
   ▪ 10 proponents including PSNR
   ▪ no winner, 8~9 proponents statistically equivalent, including PSNR!

2. Phase II test (2003):
   ▪ Restricted types of distortions (MPEG)
   ▪ Result: A few models slightly better than PSNR

3. VQEG is extending their directions:
   ▪ FR/RR/NR, Low Bit Rate
   ▪ Multimedia: video, audio and speech …
An example of non-adequacy of PSNR

The left image (distorted by additive noise) has higher visual quality than the right one (highly compressed by JPEG). At the same time PSNR analysis states that the right image quality is higher by 5.7 dB.
Visual quality metrics in image processing

For effectiveness of testing one needs to have a sample image and a quality metric that corresponds to human visual system.

Database quality is determined by correctness of image test selection, used types of distortions, and methodology to carry out for subjective image quality experiments.

Adequacy of image subjective quality evaluation is determined by a number of experiment participants and by methodology of visual quality evaluation.
Human Visual System (HVS) is a subject of intensive study during two recent decades. The study is far from completeness and the obtained knowledge is mainly fragmentary.

Human visual sensitivity varies as a function of several key image properties, such as:

- Light level
- Spatial frequency
- Color
- Local image contrast
- Eccentricity
- Temporal frequency (for video compression)
- Evenness of distortion distribution in an image
- Possibility to recover the lost information due to some robustness and adaptivity of HVS
Examples of peculiarities of HVS

Contrast Sensitivity Function (CSF)
Examples of peculiarities of HVS

An illustration to non-eccentricity of distortions: a head among coffee beans does not stretch an eye and does not influence estimation of image visual quality
Human visual sensitivity
varies as a function of several key image properties, such as:

✓ Light level
✓ Spatial frequency
✓ Color
✓ Local image contrast
✓ Eccentricity
✓ Temporal frequency

Goal:
Efficient accounting for local image contrast using a model of between-coefficient contrast masking of DCT basis functions

Masking model can be used in:
✓ Image and video compression
✓ Image filtering
✓ Digital watermarking
✓ Validation of effectiveness of image processing methods

Requirements to the model:
Images compressed (filtered or processed) with accounting the model can be visualized in unknown illumination conditions, monitor brightness, distance to the monitor, viewing angle, etc. Thus such model should operate by only some averaged parameters of image visualization
**PSNR-HVS**

- Modification of the Peak Signal-to-Noise Ratio

\[
PSNR_{\text{HVS}} = 10 \log \left( \frac{\text{max}^2}{MSE_H} \right)
\]

where

\[
MSE_H = k \sum_{i=1}^{7} \sum_{j=1}^{7} \sum_{m=1}^{8} \sum_{n=1}^{8} \left( (X[m,n]_i^c - X[m,n]_i^d)T_c[m,n] \right)^2
\]

<table>
<thead>
<tr>
<th>T[m,n]</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>1</td>
<td>1.61</td>
<td>2.34</td>
<td>2.57</td>
<td>1.61</td>
<td>1.07</td>
<td>0.64</td>
<td>0.51</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>2.14</td>
<td>2.14</td>
<td>1.84</td>
<td>1.35</td>
<td>0.99</td>
<td>0.44</td>
<td>0.43</td>
<td>0.47</td>
</tr>
<tr>
<td>3</td>
<td>1.84</td>
<td>1.98</td>
<td>1.61</td>
<td>1.07</td>
<td>0.64</td>
<td>0.45</td>
<td>0.37</td>
<td>0.46</td>
</tr>
<tr>
<td>4</td>
<td>1.84</td>
<td>1.51</td>
<td>1.17</td>
<td>0.89</td>
<td>0.51</td>
<td>0.29</td>
<td>0.32</td>
<td>0.41</td>
</tr>
<tr>
<td>5</td>
<td>1.43</td>
<td>1.17</td>
<td>0.69</td>
<td>0.46</td>
<td>0.38</td>
<td>0.24</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>6</td>
<td>1.07</td>
<td>0.73</td>
<td>0.47</td>
<td>0.40</td>
<td>0.32</td>
<td>0.25</td>
<td>0.23</td>
<td>0.28</td>
</tr>
<tr>
<td>7</td>
<td>0.52</td>
<td>0.40</td>
<td>0.33</td>
<td>0.29</td>
<td>0.25</td>
<td>0.22</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>8</td>
<td>0.38</td>
<td>0.28</td>
<td>0.27</td>
<td>0.26</td>
<td>0.23</td>
<td>0.26</td>
<td>0.25</td>
<td>0.26</td>
</tr>
</tbody>
</table>
PSNR-HVS-M

**PSNR\textsubscript{HVSM}**

- **Problem:** masking effect too strong in blocks including edges.

- **Solution:** masking effect reduction proportionally to the local variances, $\text{var}(.)$ in the sub-blocks $B_1$, $B_2$, $B_3$ and $B_4$

\[
E_m(B) = \frac{E_w(B) \cdot \delta(B)}{16}
\]

where

\[
\delta(B) = \frac{\text{var}(B_1) + \text{var}(B_2) + \text{var}(B_3) + \text{var}(B_4)}{4 \cdot \text{var}(B)}
\]
PSNR-HVS-M

The maximum masking effect coefficient is defined as:

$$E_{\text{max}} = \max(E_m(X_o), E_m(X_d))$$

The $MSE_H$ is then:

$$MSE_H = K \sum_{i=1}^{l-7} \sum_{j=1}^{J-7} \sum_{m=1}^{8} \sum_{n=1}^{8} \left( X[m,n]_{i,j}^o - X[m,n]_{i,j}^d \right)^2 T_c[m,n]$$

where:

$$X[m,n]_{i,j}^o - X[m,n]_{i,j}^d = \begin{cases} 
X_{i,j}^o - X_{i,j}^d, & i = j = 0 \\
0, & |X_{i,j}^o - X_{i,j}^d| < \frac{E_{\text{norm}}}{C_{i,j}} \\
X_{i,j}^o - X_{i,j}^d - \frac{E_{\text{norm}}}{C_{i,j}}, & X_{i,j}^o - X_{i,j}^d > \frac{E_{\text{norm}}}{C_{i,j}} \\
X_{i,j}^o - X_{i,j}^d + \frac{E_{\text{norm}}}{C_{i,j}}, & \text{otherwise}
\end{cases}$$
**PSNR-HVS and PSNR-HVS-M**


**Flow-chart of PSNR-HVS-M calculation**

1. Block 8x8 of original image
2. DCT of difference between pixel values
3. Reduction by value of contrast masking
4. MSE calculation of the block
Structural Similarity (SSIM) Index in Image Space

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

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

\[ s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \]

\[ SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) \]


[Wang et al., *IEEE Trans. Image Processing*, ’04]
Many tasks for which we need to evaluate visual quality of obtained images having a sample (reference) image:

- Design of image denoising methods;
- Design of image and video compression methods;
- Design of digital watermarking techniques;
- Effectiveness evaluation of other image processing techniques.

Many different quality metrics have been designed for image quality evaluation. Classical metrics, MSE and PSNR, do not take into account any peculiarity of human visual system (HVS). Therefore, their use often leads to non adequate evaluation of processed image visual quality.
Requirements to a test image database

Image database should reflect peculiarities of HVS and should contain images non-trivial for visual quality evaluation in order to effectively retrieve all advantages and drawbacks of all tested quality metrics.

Taking this into account, it is possible to set the following requirements to the test image database:

• It should include images with considerably different characteristics: percentage of homogeneous regions, details and textures, various texture characteristics, etc.
• For each peculiarity of HVS, the database should contain a distortion type that allows estimating the influence degree of this peculiarity on image visual quality
• Images in the database should not be too simple for visual quality estimation. 1) A number of distortion levels should not be too large. 2) A number of situations when all metrics evidence in favor of a given image should not be large.
• It is desirable that the database contains image distortions typical for practice that originate due to compression, denoising, data transmission errors, etc.
Examples of undesirable situations

Too simple situations, when all quality metrics will evidence in favor of better quality of the right image. Verification of quality metrics for such test sets will produce over estimated values of their effectiveness (correspondence to HVS)
Examples of undesirable situations

Next undesirable situation when both images contain the same type of distortions of different level. In this case all quality metrics will evidence in favor of better quality of the right image. This will decrease accuracy of verification results.
# Details of TID2008 content

<table>
<thead>
<tr>
<th>№</th>
<th>Type of distortion (4 levels for each distortion)</th>
<th>Correspondence to practice situation</th>
<th>Accounting of HVS peculiarities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Additive Gaussian noise</td>
<td>Image acquisition</td>
<td>Adaptivity, robustness</td>
</tr>
<tr>
<td>2</td>
<td>Additive noise in color components is more intensive than additive noise in the luminance component</td>
<td>Image acquisition</td>
<td>Color sensitivity</td>
</tr>
<tr>
<td>3</td>
<td>Spatially correlated noise</td>
<td>Digital photography</td>
<td>Spatial frequency sensitivity</td>
</tr>
<tr>
<td>4</td>
<td>Masked noise</td>
<td>Image compression, watermarking</td>
<td>Local contrast sensitivity</td>
</tr>
<tr>
<td>5</td>
<td>High frequency noise</td>
<td>Image compression, watermarking</td>
<td>Spatial frequency sensitivity</td>
</tr>
<tr>
<td>6</td>
<td>Impulse uniform noise</td>
<td>Image acquisition</td>
<td>Robustness</td>
</tr>
<tr>
<td>7</td>
<td>Quantization noise</td>
<td>Image registration, gamma correction</td>
<td>Color, local contrast, spatial frequency</td>
</tr>
<tr>
<td>8</td>
<td>Gaussian blurring</td>
<td>Image registration, image compression</td>
<td>Spatial frequency sensitivity</td>
</tr>
<tr>
<td>9</td>
<td>DCT 3D Denoising</td>
<td>Image denoising</td>
<td>Spatial frequency, local contrast</td>
</tr>
<tr>
<td>10</td>
<td>JPEG compression</td>
<td>JPEG compression</td>
<td>Color, spatial frequency sensitivity</td>
</tr>
<tr>
<td>11</td>
<td>JPEG2000 compression</td>
<td>JPEG2000 compression</td>
<td>spatial frequency sensitivity</td>
</tr>
<tr>
<td>12</td>
<td>JPEG transmission errors</td>
<td>Data transmission</td>
<td>Eccentricity</td>
</tr>
<tr>
<td>13</td>
<td>JPEG2000 transmission errors</td>
<td>Data transmission</td>
<td>Eccentricity</td>
</tr>
<tr>
<td>14</td>
<td>Non eccentricity pattern noise</td>
<td>Image compression, watermarking</td>
<td>Eccentricity</td>
</tr>
<tr>
<td>15</td>
<td>Local block-wise distortions of different intensity</td>
<td>Image acquisition</td>
<td>Evenness of distortions</td>
</tr>
<tr>
<td>16</td>
<td>Mean shift (intensity shift)</td>
<td>Image acquisition</td>
<td>Light level sensitivity</td>
</tr>
<tr>
<td>17</td>
<td>Contrast change</td>
<td>Image acquisition, gamma correction</td>
<td>Light level, local contrast sensitivity</td>
</tr>
</tbody>
</table>
Examples of some types of distortions: the result of 3D DCT denoising

Additive Gaussian noise

PSNR(Y)=25.81 dB
PSNR-HVS-M=29.52 dB
SSIM=0.69

Result of 3D DCT denoising

PSNR(Y)=25.97 dB
PSNR-HVS-M=24.95 dB
SSIM=0.65
Examples of some types of distortions: non-eccentricity distortions

Sample image

Distorted image, PSNR=27 dB
Examples of some types of distortions: block-wise distortions

PSNR=26.79 dB
PSNR-HVS-M=24.41 dB
SSIM=0.95

PSNR=26.11 dB
PSNR-HVS-M=25.00 dB
SSIM=0.99

For this distortion type we check hypothesis that after some threshold level human eye does not react to distortion intensity but it reacts to total distorted area.
Set of test images

First 24 images are fragments (512x384 pixels) of Kodak test image set
Set of test images

25th artificial image is added to provide the possibility of quality metrics verification for such images quality evaluation
Basic principles of experiments carrying out and results control

Main problem – too large time needed for carrying out full sorting of the image set that contains 1700 images.

It seems more reasonable to carry out full sorting of image set for each sample image separately (25 times for 68 images). Such approach allows decrease the total experiments time by up to 5 times. This also allows dividing the experiment into some separate independent experiments. In this manner the time of a separate experiment decreases approximately by 100 times and becomes about 15-25 minutes. This allows attracting a large number of experiment participants. Then due to linearity of Spearman and Kendall correlations it becomes possible to calculate this correlations as averaged values for all 25 sample images.

In experiments pairs of images at the screen are changed with some delay in order to avoid influence of temporal frequencies on HVS.

After carrying out of each experiment, Spearman correlation factor is evaluated between a given participant and averaged data. The result is considered reliable if this correlation factor is over 0.5.

A participant should not spend more than 2-3 seconds for each comparison of visual quality of a pair of images. 100-150 participants (each participant processes 2-3 sample images) are needed.
What does such database creation allow

Such database creation allows:
• To obtain possibility of reliable verification of existing and designed quality metrics;
• To design new quality metrics;
• To design new image processing methods for image compression;
  watermarking;
  denoising;
  deblurring;
  etc.
Description:

• This homework is aimed to carry a subjective image quality evaluation. Currently, it is the most widely used method to study the influence of various image distortion such as noise, or results of applying different image processing methods to a human visual system (HVS).

• In this homework, you are requested to do 2-3 experiments. The deadline is 4 March, 2013.

• The following shows instructions for one experiment. Please read carefully before doing an experiment.

  Sit in front of the screen. Screen resolution should be not less than 1152x864. You can use your own computer connected to TUT internet or computer in Signal Processing Laboratory, e.g. TC407, TC303. Wear any viewing device you need. Be at comfortable seating distance from the monitor. Carry out experiments in good illumination conditions.

Click the link: [http://tid2013.cs.tut.fi/](http://tid2013.cs.tut.fi/) , you will see
Quality assessment: homework

Subjective experiments for TIE043.

Your name: ____________________________ Image set: 16

Start new experiment

Distorted Image 1

Distorted Image 2

Original Image
Quality assessment: homework

Step 1: Insert your student number in the box 'Your name' with ‘-A’, ‘-B’ and ‘-C’ for your first, second and third image set experiment, respectively.

For example, if your student number is 123456,
123456-A: indicate your first experiment,
123456-B: indicate your second experiment,
123456-C: indicate your third experiment.

Step 2: Keep or change (by typing a number from 1 to 25) an automatically randomly selected image set in the box 'Image set'.

Step 3: The experiment now can be started by clicking the button 'Start new experiment':
   a. In the bottom part you will be shown the original image.
   b. In the upper part there are two distorted images.

Step 4: Click on the distorted image you think is more similar to the original one.

Please use 2-3 seconds (not more) for each selection. If the quality of the two distorted images is comparable please click on any of them.
During one experiment you will need to perform 540 comparisons. A bar in the upper part indicates a percentage of already performed comparisons.

Total duration of one experiment is about 30 minutes. Please have at least ten minutes rest to carry out the next experiment.

When an experiment is completed you’ll see the message: "The experiment is completed". After pressing "OK" button results of the experiment will be transferred to a database.

For any questions, please contact with Lina Jin, TE413, lina.jin@tut.fi.