A Simple Model for Assessing Perceptual Image Quality

Rafael Reisenhofer

AG Computational Data Analysis, Universität Bremen

Colloquium
Institute of Cognitive Science, Universität Osnabrück
25 October 2017
Outline

Perceptual Image Quality Assessment

The Haar Wavelet-based Perceptual Similarity Index (HaarPSI)

Experimental Validation of the HaarPSI

Human Perception of Image Quality
Basic Applied Harmonic Analysis Pipeline

Things we do:

- Find optimal transforms for specific classes of data and prove their optimality.
- Develop transform-based algorithms to obtain desired results.
- Formalize and prove properties of transforms and algorithms.
Measuring Perceptual Distortion is Difficult

Additive Gaussian Noise

MSE: 820.7282

Block Distortions

MSE: 820.7203

We want to define an image similarity measure that reflects the degree of similarity experienced by a human observer.
Full Reference Image Quality Assessment

Reference $f_1$  
Distorted $f_2$

We aim to define a perceptual similarity measure

$$\text{HaarPSI}: \ell_2(\mathbb{Z}^2) \times \ell_2(\mathbb{Z}^2) \rightarrow [0, 1]$$

that correctly predicts the human perception of similarity between a reference image and its distorted version.
Some Well-Known Image Similarity Measures

- PSNR: Peak Signal-to-Noise Ratio
- MS-SSIM: Multi-Scale Structural Similarity Index (Z. Wang, E. P. Simoncelli & A. C. Bovik; 2003)
- VIF: Visual Information Fidelity (H. R. Sheikh & A. C. Bovik; 2006)
- MAD: Most Apparent Distortion (E. C. Larson & D. M. Chandler; 2010)
- GSM: Gradient Similarity-based Metric (A. Liu, W. Lin, & M. Narwaria; 2012)
- SR-SIM: Spectral Residual Similarity Index (L. Zhang & H. Li; 2012)
- VSI: Visual Saliency Index (L. Zhang, Y. Shen, & H. Li; 2014)
- NLP: Normalized Laplacian Pyramid (V. Laparra, J. Ballé, A. Berardino & E. P. Simoncelli; 2016)
Towards the HaarPSI

Basic principle: Locally compute features for both images and compare them with a scalar similarity measure

\[ S(a, b, C) = \frac{2ab + C}{a^2 + b^2 + C}, \]

with a parameter \( C > 0 \).

Images taken from the CSIQ database
Guiding Principles

- Incorporate neurophysiological principles on a basic functional level
  - Response to singularities (edges/ridges)
  - Orientation selectivity
  - Frequency selectivity
- Activation function
- Visual saliency (some parts of an image are more important than others)
Basic Approach

For two images $f_1$ and $f_2$, compute a weighted average using

- a local similarity map $HS_{f_1,f_2}^{(k)}$
- a weight map $W_{f_1,f_2}^{(k)}$
- and a logistic activation function $l_\alpha(x) = \frac{1}{1 + e^{-\alpha x}}$,

$$\text{HaarPSI}_{f_1,f_2} \approx \frac{\sum_x l_\alpha(HS[x]) \cdot W[x]}{\sum_x W[x]}.$$
Discrete Haar Wavelet Filters

Using the 1D filter coefficients

\[ h_{1D}^1 = \frac{[1, 1]}{\sqrt{2}}, \]
\[ g_{1D}^1 = \frac{[-1, 1]}{\sqrt{2}}, \]
\[ g_{jD} = h_{1D}^1 * (g_{j-1D})^\uparrow^2, \]
\[ h_{jD} = h_{1D}^1 * (h_{j-1D})^\uparrow^2, \]

horizontal and vertical 2D Haar filters

\[ g_{j}^{(1)} = g_{jD}^1 \otimes h_{jD}^1, \]
\[ g_{j}^{(2)} = h_{jD}^1 \otimes g_{jD}^1, \]

are constructed for a scale \( j \in \mathbb{N} \).

For \( h_{1D}^1 \) and \( g_{1D}^1 \), any pair of wavelet lowpass/highpass filters could be used!
HaarPSI Similarity Maps

\[
\text{HS}^{(k)}_{f_1, f_2}[x] = l_\alpha \left( \frac{1}{2} \sum_{j=1}^{2} S \left( |(g_j^{(k)} * f_1)[x]|, |(g_j^{(k)} * f_2)[x]|, C \right) \right)
\]

- Features are constructed by convolving the images with discrete Haar wavelet filters.
- Only the two finest scales of the transform are used.
- The only free parameters are \( \alpha \) and \( C \).
HaarPSI Weight Maps

\[ \mathbf{W}_f^{(k)}[x] = \left| (g_3^{(k)} * f)[x] \right|, \]

- Weights are also obtained by convolving an image with discrete Haar wavelet filters.
- Only the coarsest Haar filters are used.
The Haar Wavelet-Based Perceptual Similarity Index

Combining the similarity maps $\mathbf{HS}_{f_1,f_2}^{(k)}$ and the weight maps $\mathbf{W}_f^{(k)}$, the HaarPSI for two grayscale images $f_1, f_2$ is given by the weighted average

$$\text{HaarPSI}_{f_1,f_2} = l^{-1}_\alpha \left( \frac{\sum_x \sum_{k=1}^2 \mathbf{HS}_{f_1,f_2}^{(k)}[x] \cdot \mathbf{W}_f^{(k)}[x]}{\sum_x \sum_{k=1}^2 \mathbf{W}_f^{(k)}[x]} \right)^2,$$

with

$$\mathbf{W}_{f_1,f_2}[x] = \max(\mathbf{W}_f^{(k)}[x], \mathbf{W}_f^{(k)}[x]).$$

- The HaarPSI can be generalized to color images in the YIQ color space
Benchmark Databases

The consistency of HaarPSI with the human perception of image quality was evaluated on four large publicly available benchmark databases of quality-annotated color images:

- LIVE (H. R. Sheikh, Z. Wang, L. Cormack & A. C. Bovik)
  - 29 reference images, 5 distortion types, 779 distorted images
- TID 2008 (N. Ponomarenko et al.; 2009)
  - 25 reference images, 17 distortion types and 1700 distorted images
- TID 2013 (N. Ponomarenko et al.; 2015)
  - 25 reference images, 24 distortion types and 3000 distorted images
-CSIQ (E. C. Larson & D. M. Chandler; 2010)
  - 30 reference images, 6 distortion types and 866 distorted images
The Parameters \( C \) and \( \alpha \)

The two free parameters parameters were optimized on randomly chosen subsets of four large publicly available databases with respect to the mean Spearman rank order correlation coefficient (SROCC):

\[
C = 30 \quad \text{and} \quad \alpha = 4.2
\]

\[
\text{HS}_{f_1, f_2}^{(k)}[x] = l_\alpha \left( \frac{1}{2} \sum_{j=1}^{2} S \left( \left| (g_j^{(k)} \ast f_1)[x] \right|, \left| (g_j^{(k)} \ast f_2)[x] \right|, C \right) \right)
\]

\[
l_\alpha(x) = \frac{1}{1 + e^{-\alpha x}}
\]

\[
S(a, b, C) = \frac{2ab + C}{a^2 + b^2 + C}
\]
Correlation with Human Opinion Scores

- Fitted four parameter logistic function
- Gaussian noise
- Gaussian noise (color intensive)
- Spatially correlated noise
- Masked noise
- High frequency noise
- Impulse noise
- Quantization noise
- Gaussian blur
- Image denoising
- JPEG compression
- JPEG2000 compression
- JPEG transmission error
- JPEG2000 transmission error
- Non eccentricity pattern noise
- Block-wise distortions
- Mean shift
- Contrast change
- Change of color saturation
- Multiplicative Gaussian noise
- Comfort noise
- Lossy compression of noisy images
- Image color quantization w dither
- Chromatic aberrations
- Sparse sampling and reconstruction
- Gaussian pink noise
- Fast fading

LIVE

TID 2008

TID 2013

CSIQ
## Comparison of Different IQA Metrics

### Grayscale Images

<table>
<thead>
<tr>
<th></th>
<th>PSNR</th>
<th>VIF</th>
<th>SSIM</th>
<th>MS-SSIM</th>
<th>GSM</th>
<th>MAD</th>
<th>SR-SIM</th>
<th>FSIM</th>
<th>VSI</th>
<th>HaarPSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIVE</td>
<td>0.8756</td>
<td>0.9636</td>
<td>0.9479</td>
<td>0.9513</td>
<td>0.9561</td>
<td>0.9672</td>
<td>0.9619</td>
<td>0.9634</td>
<td>0.9534</td>
<td><strong>0.9690</strong></td>
</tr>
<tr>
<td>TID2008</td>
<td>0.5531</td>
<td>0.7491</td>
<td>0.7749</td>
<td>0.8542</td>
<td>0.8504</td>
<td>0.8340</td>
<td>0.8913</td>
<td>0.8804</td>
<td>0.8830</td>
<td><strong>0.9043</strong></td>
</tr>
<tr>
<td>TID2013</td>
<td>0.6394</td>
<td>0.6769</td>
<td>0.7417</td>
<td>0.7859</td>
<td><strong>0.7946</strong></td>
<td>0.7807</td>
<td>0.8075</td>
<td>0.8022</td>
<td>0.8048</td>
<td><strong>0.8094</strong></td>
</tr>
<tr>
<td>CSIQ</td>
<td>0.8058</td>
<td>0.9195</td>
<td>0.8756</td>
<td>0.9133</td>
<td>0.9108</td>
<td>0.9466</td>
<td>0.9319</td>
<td>0.9242</td>
<td>0.9372</td>
<td><strong>0.9546</strong></td>
</tr>
</tbody>
</table>

### Color Images

<table>
<thead>
<tr>
<th></th>
<th>PSNR</th>
<th>VIF</th>
<th>SSIM</th>
<th>MS-SSIM</th>
<th>GSM</th>
<th>MAD</th>
<th>SR-SIM</th>
<th>FSIM</th>
<th>VSI</th>
<th>HaarPSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIVE</td>
<td>0.8756</td>
<td>0.9636</td>
<td>0.9479</td>
<td>0.9513</td>
<td>0.9561</td>
<td>0.9672</td>
<td>0.9619</td>
<td>0.9645</td>
<td>0.9524</td>
<td><strong>0.9683</strong></td>
</tr>
<tr>
<td>TID2008</td>
<td>0.5531</td>
<td>0.7491</td>
<td>0.7749</td>
<td>0.8542</td>
<td>0.8504</td>
<td>0.8340</td>
<td>0.8913</td>
<td>0.8840</td>
<td>0.8979</td>
<td><strong>0.9097</strong></td>
</tr>
<tr>
<td>TID2013</td>
<td>0.6394</td>
<td>0.6769</td>
<td>0.7417</td>
<td>0.7859</td>
<td><strong>0.7946</strong></td>
<td>0.7807</td>
<td>0.8075</td>
<td>0.8510</td>
<td>0.8965</td>
<td><strong>0.8965</strong></td>
</tr>
<tr>
<td>CSIQ</td>
<td>0.8058</td>
<td>0.9195</td>
<td>0.8756</td>
<td>0.9133</td>
<td>0.9108</td>
<td>0.9466</td>
<td>0.9319</td>
<td>0.9310</td>
<td>0.9423</td>
<td><strong>0.9604</strong></td>
</tr>
</tbody>
</table>

Lower correlation than HaarPSI. The difference is statistically significant with \( p < 0.05 \).

Higher correlation than HaarPSI. The difference is statistically significant with \( p < 0.05 \).

The highest correlation in each row is written in **boldface**.

### Spearman Rank Order Correlations (SROCC) of IQA Metrics With Human Mean Opinion Scores

17/23
Intermediate Summary

- Measuring perceptual distortion is difficult!

- A simple approach implementing functional properties of the HVS on a very basic level suffices to achieve state of the art correlations with human opinion scores.

- Can the HaarPSI and considering human opinion scores as a baselines help us understand mechanisms of the human perception of image quality?
The Parameters $C$ and $\alpha$

\[ S(x, y, C) = \frac{2xy + C}{x^2 + y^2 + C} \]

\[ l_\alpha(x) = \frac{1}{1 + e^{-\alpha x}} \]

- Increasing $C$ decreases the sensitivity of the HaarPSI to changes in the high-frequency components measured by the similarity maps $\text{HS}_{f_1,f_2}^{(1,2)}$ relative to the weights $\textbf{W}_f^{(1,2)}$.
- The effect of the parameter $\alpha$ on the HaarPSI is qualitatively similar when it is approaching zero.
- For larger choices of $\alpha$, the function $l_\alpha(\cdot)$ is mimicking the behavior of a thresholding operator in the sense that only severe changes in the high-frequency components will have a significant effect.
Analysis of Specific Distortions

SROCC as functions of $C$ and $\alpha$. 
Eigendistortions of Image Representations

- How well can more complex hierarchical models (e.g. deep neural networks) explain the underlying mechanisms of the human perception of image similarity?

- Which other ways exist of comparing the validity of two models?

- For a image representation $f$ and an image $x$, Berardino et al. recently proposed to consider the extremal eigenvectors of the Fisher information matrix

$$J(x) = \frac{\partial f}{\partial x}^T \frac{\partial f}{\partial x},$$

where $\frac{\partial f}{\partial x}$ denotes the Jacobian matrix.

Berardino, Ballé, Laparra and Simoncelli; 2017
Final Remarks

- The HaarPSI incorporates neurophysiological principles like sensitivity to singularities, orientation and frequency on a very basic, functional level. It only uses six discrete Haar wavelet filters.

- The HaarPSI has high correlations with human opinion scores on four large databases.

- It is surprising that Haar filters outperform other wavelet filters.

- A principle that is currently missing from the HaarPSI is divisive normalization.

- When considering extremal distortions, shallow models inspired by neurobiology seem to better match the human sensitivity than generic deep representations.
References


www.haarpsi.org

Thank you for your time!
Using Other Wavelet Filters

<table>
<thead>
<tr>
<th></th>
<th>Daub2PSI</th>
<th>Daub4PSI</th>
<th>Sym4PSI</th>
<th>CDFPSI</th>
<th>Coif1PSI</th>
<th>HaarPSI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grayscale Images</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIVE</td>
<td>0.9593</td>
<td>0.9507</td>
<td>0.9527</td>
<td>0.9507</td>
<td>0.9579</td>
<td><strong>0.9675</strong></td>
</tr>
<tr>
<td>TID 2008</td>
<td>0.8852</td>
<td>0.8539</td>
<td>0.8746</td>
<td>0.8742</td>
<td>0.8828</td>
<td><strong>0.9042</strong></td>
</tr>
<tr>
<td>TID 2013</td>
<td>0.8022</td>
<td>0.7819</td>
<td>0.7962</td>
<td>0.7946</td>
<td>0.8019</td>
<td><strong>0.8129</strong></td>
</tr>
<tr>
<td>CSIQ</td>
<td>0.9492</td>
<td>0.9453</td>
<td>0.9453</td>
<td>0.9413</td>
<td>0.9482</td>
<td><strong>0.9525</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Daub2PSI</th>
<th>Daub4PSI</th>
<th>Sym4PSI</th>
<th>CDFPSI</th>
<th>Coif1PSI</th>
<th>HaarPSI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Color Images</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIVE</td>
<td>0.9623</td>
<td>0.9572</td>
<td>0.9601</td>
<td>0.9594</td>
<td>0.9618</td>
<td><strong>0.9685</strong></td>
</tr>
<tr>
<td>TID 2008</td>
<td>0.8896</td>
<td>0.8612</td>
<td>0.8803</td>
<td>0.8822</td>
<td>0.8864</td>
<td><strong>0.9079</strong></td>
</tr>
<tr>
<td>TID 2013</td>
<td>0.8700</td>
<td>0.8569</td>
<td>0.8675</td>
<td>0.8672</td>
<td>0.8719</td>
<td><strong>0.8791</strong></td>
</tr>
<tr>
<td>CSIQ</td>
<td><strong>0.9580</strong></td>
<td>0.9548</td>
<td>0.9561</td>
<td>0.9564</td>
<td>0.9561</td>
<td>0.9566</td>
</tr>
</tbody>
</table>
# Overall Performance and Execution Time

Comparison of mean Spearman rank order correlations and execution times

<table>
<thead>
<tr>
<th></th>
<th>Color Images</th>
<th>Grayscale Images</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SROCC</td>
<td>Time (ms)</td>
<td>SROCC</td>
<td>Time (ms)</td>
</tr>
<tr>
<td>HaarPSI</td>
<td>0.9279</td>
<td>24</td>
<td>0.9093</td>
<td>10</td>
</tr>
<tr>
<td>VSI</td>
<td>0.9223</td>
<td>79</td>
<td>0.8946</td>
<td>80</td>
</tr>
<tr>
<td>FSIM</td>
<td>0.9076</td>
<td>142</td>
<td>0.8925</td>
<td>121</td>
</tr>
<tr>
<td>SRSIM</td>
<td>0.8982</td>
<td>10</td>
<td>0.8982</td>
<td>10</td>
</tr>
<tr>
<td>MAD</td>
<td>0.8821</td>
<td>892</td>
<td>0.8821</td>
<td>891</td>
</tr>
<tr>
<td>GSM</td>
<td>0.8780</td>
<td>8</td>
<td>0.8780</td>
<td>7</td>
</tr>
<tr>
<td>MSSSIM</td>
<td>0.8762</td>
<td>30</td>
<td>0.8762</td>
<td>24</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.8350</td>
<td>6</td>
<td>0.8350</td>
<td>5</td>
</tr>
<tr>
<td>VIF</td>
<td>0.8273</td>
<td>459</td>
<td>0.8273</td>
<td>453</td>
</tr>
<tr>
<td>PSNR</td>
<td>0.7185</td>
<td>2</td>
<td>0.7185</td>
<td>1</td>
</tr>
</tbody>
</table>
HaarPSI Examples

\[ f_1 \quad \text{and} \quad f_2 \ (\text{JPEG compression}) \]

[Images of two similar buildings, one of which is JPEG compressed]

\[ \text{HaarPSI}_{f_1,f_2} = 0.9439 \]

Images taken from the TID 2013 database
HaarPSI Examples

\[ f_1 \quad f_2 \text{ (JPEG compression)} \]

\[
\text{HaarPSI}_{f_1,f_2} = 0.6218
\]

Images taken from the TID 2013 database
HaarPSI Examples

\[ f_1 \quad f_2 \text{ (JPEG compression)} \]

\[
\text{HaarPSI}_{f_1,f_2} = 0.4272
\]

Images taken from the TID 2013 database
HaarPSI Examples

\[ f_1, f_2 = 0.2166 \]

Images taken from the TID 2013 database