# A Haar Wavelet-Based Perceptual Similarity Index for Image Quality Assessment

Rafael Reisenhofer* Sebastian Bosse† Gitta Kutyniok‡ Thomas Wiegand§

**Abstract**

In most practical situations, images and videos can neither be compressed nor transmitted without introducing distortions that will eventually be perceived by a human observer. Vice versa, most applications of image and video restoration techniques, such as inpainting or denoising, aim to enhance the quality of experience of human viewers. Correctly predicting the similarity of an image with an undistorted reference image, as subjectively experienced by a human viewer, can thus lead to significant improvements in any transmission, compression, or restoration system. This paper introduces the Haar wavelet-based perceptual similarity index (HaarPSI), a novel and easy-to-compute similarity measure for full reference image quality assessment. HaarPSI utilizes the coefficients obtained from a Haar wavelet decomposition to assess local similarities between two images, as well as the relative importance of image areas. The consistency of HaarPSI with human quality of experience was validated on four large benchmark databases containing several thousands of differently distorted images. On these databases, HaarPSI achieves higher correlations with human opinion scores than state-of-the-art full reference similarity measures like the structural similarity index (SSIM), the feature similarity index (FSIM), and the visual saliency-based index (VSI). Along with the simple computational structure and the short execution time, these promising experimental results suggest a high applicability of HaarPSI in real world tasks.

## 1 Introduction

Today, digital images and videos have become an ubiquitous means of representing and communicating information. Modern hand-held devices such as mobile phones are omnipresent in our daily lives and are typically fully equipped with hardware and software to capture, store, send, and display visual signals. Digital video broadcasts allow for the simultaneous transmission of a huge amount of different channels and internet-based streaming of video signals continues to be on the rise.

*R. Reisenhofer is with the Working Group Computational Data Analysis, Universität Bremen, Fachbereich 3, Postfach 330440, 28334 Bremen, Germany (e-mail: reisenhofer@math.uni-bremen.de).

†S. Bosse is with the Fraunhofer Heinrich Hertz Institute (Fraunhofer HHI), 10587 Berlin, Germany (e-mail: sebastian.bosse@hhi.fraunhofer.de).

‡G. Kutyniok is with the Department of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany (e-mail: kutyniok@math.tu-berlin.de) and acknowledges support by the Einstein Foundation Berlin, the Einstein Center for Mathematics Berlin (ECMath), the European Commission-Project DEDALE (contract no. 665044) within the H2020 Framework Program, DFG Grant KU 1446/18, DFG-SPP 1798 Grants KU 1446/21 and KU 1446/23, the DFG Collaborative Research Center TRR 109 Discretization in Geometry and Dynamics, and by the DFG Research Center Matheon Mathematics for Key Technologies in Berlin.

§T. Wiegand is with the Fraunhofer Heinrich Hertz Institute (Fraunhofer HHI), 10587 Berlin, Germany, and with the Image Communication Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany (e-mail: twiegand@ieee.org).

In most image and video transmission systems, distortions are introduced to the transmitted signal. For most applications, the ultimate receiver is the human visual system. Thus, besides objective factors such as the packet loss rate or the throughput, the Quality of Experience (QoE) in such systems is highly affected by the subjectively perceived quality of the received multimedia signal. The same is true for image and video restoration techniques, such as denoising or inpainting, which are usually applied to enhance the experience of human viewers. Correctly predicting the subjectively experienced similarity of an image with an undistorted reference image, by applying a numerical image quality measure, can hence lead to a significant improvement of the QoE of a transmission, compression, or restoration system.

Image quality assessment methods can be differentiated by how much information about the undistorted reference image is used for estimating the perceived quality of the corresponding distorted image: While full reference (FR) image quality assessment approaches assume and exploit the full knowledge of the reference image, no reference (NR) image quality assessment methods do not rely on any information regarding the reference image (although it is typically assumed to be a natural image). Reduced reference (RR) quality assessment approaches exist in the middle of this spectrum and require a small set of features derived from the reference image for estimating the similarity with the considered distorted image.

In transmission systems, the complete reference image is typically available to the sender, while towards the receiver's end, transmitting information about the reference image becomes increasingly costly. This leads to different application domains for FR, NR, and RR approaches: FR methods are useful at the beginning of a transmission process, e.g. for controlling an encoding procedure, during which the original image is fully available. NR methods, on the other hand, can also be helpful on the receiver's side, where only the distorted, transmitted signal is available, e.g. for controlling deblurring algorithms, or other forms of post-processing. Finally, RR methods can be applied throughout the whole transmission process, given that the features of the reference image required for a good quality assessment only use a small amount of bandwidth.

In addition to practical applications, research on image and video quality assessment might also shed light on the underlying mechanisms of human visual perception and can provide a framework for the evaluation of computational models of human vision.

Usually, image quality measures are evaluated and compared by considering rank order correlations with so-called mean opinion scores (MOS), that have previously been experimentally obtained for large databases of distorted images. The MOS is a numerical value assigned to images in psychophysical tests during which participants rate the subjectively perceived quality of said images. The MOS is typically considered to be the ground truth for the perceived quality of a distorted image.

This work introduces the Haar wavelet-based perceptual similarity index (HaarPSI), a novel and computationally inexpensive algorithm yielding FR image quality assessments. The basic idea of HaarPSI is to judge the similarity between two images by considering their respective Haar wavelet representations. The magnitude responses on the two scales of the wavelet transform associated with the highest frequencies are used to compute local similarities. Furthermore, the relative importance of thus obtained local (dis)similarities is assessed by a weight function that takes the sum over all four scales of the Haar wavelet transform at a given location. Combining the similarity map with the weight function yields the Haar similarity index, which is formally defined in (11) and (13) in Section 3. In Section 4, we evaluate the consistency of HaarPSI with the human quality of experience and compare its performance to state-of-the-art similarity measures like SSIM [1], FSIM [2], and VSI [3]. As depicted in Tables 1 and 2, HaarPSI achieves higher correlations with human opinion scores than all other considered FR quality metrics in all test cases except one, where it only comes second to VSI. In addition,

HaarPSI is significantly faster than the metrics yielding the second and third highest correlations with human opinion scores, namely VSI and FSIM.

As a final introductory remark, we would like to note that it is both convenient and surprising, that the very promising experimental results of HaarPSI reported in Section 4 are solely based on the responses of Haar filters, which are arguably the simplest and computationally most efficient wavelet filters existing. The results of a more elaborate analysis of the applicability of other wavelet filters in the similarity measure defined in Section 3 can be found in Table 4.

## 2 Previous Work

The simplest and probably still most popular FR image quality metric is the mean squared error (MSE). The MSE is defined as the average of the squared intensity difference of a distorted and a reference image (i.e., up to a factor, the squared $\ell_2$-norm of the difference). The peak signal-to-noise ratio (PSNR) is a related metric that calculates the ratio between the squared maximal intensity and the MSE and expresses it in the logarithmic decibel scale. MSE and PSNR are popular as they are cheap to compute and applicable to optimization in a straight-forward manner. However, neither MSE nor PSNR correlate well with the human perception of visual quality, as indicated by Table 1. This has led to the construction of various image quality metrics that aim for a better conformance with the human perception of image quality and image similarity.

More sophisticated approaches towards perceptually accurate image quality assessments (IQA) typically follow one of three strategies. *Bottom-up* approaches explicitly model various processing mechanism of the human visual system (HVS), such as masking effects [4], contrast sensitivity [5], or just-noticeable-distortion [6, 7], in order to assess the perceived quality of images. For instance, the adaptivity of the HVS to the magnitude of distortions is modeled explicitly by most apparent distortion (MAD) [8], in order to apply two different assessment strategies for supra- and super-threshold distortions.

However, the newly proposed Haar similiarty index as well as most image quality metrics developed recently, follow a *top-down* approach. There, general functional properties of the HVS (considered as a black box) are assumed, in order to identify and to exploit image features corresponding to the perceived quality. Prominent examples are the structural similarity index (SSIM) [1], visual information fidelity (VIF) [9], gradient similarity measure (GSM) [10], spectral residual based similarity (SR-SIM) [11], and the visual saliency-induced index (VSI) [3]. SSIM [1] tries taking into account the sensitivity of the human visual system towards structural information. This is done by pooling three complementary components, namely luminance similarity (comparing local mean luminance values), contrast similarity (comparing local variances) and structural similarity, which is defined as the local covariance between the reference image and its perturbed counterpart. Although being criticized [12], it is highly cited and among the most popular image quality assessment metrics. SSIM was generalized for a multi-scale setting by the multi-scale structural similarity index (MS-SSIM) [13]. Visual information fidelity (VIF) [9] considers the mutual information shared by a reference image and a distorted image, which are both expressed in an image model defined in the wavelet domain. Eventually, the mutual information is related to the subjectively perceived image quality. Following the basic framework of combining complementary feature maps introduced in [1], changes in contrast and structure are captured by considering local gradients in [10], while the squared difference in pixel values between the reference image and the distorted image is used to measure luminance variations. Additionally, masking effects are estimated, based on the local gradient magnitude of the reference image and incorporated when the two feature maps are combined.

A combination of two feature maps is also applied successfully by the feature similarity index (FSIM) [2] and will be discussed in more detail later in this section. Spectral residual-based similarity (SR-SIM) [11] takes into account changes in the local horizontal and vertical gradient magnitudes. Additionally, it incorporates changes in a spectral residual-based visual saliency estimate. The visual saliency-induced index (VSI) [3] follows the same line as SR-SIM by combining similarities in the gradient magnitude and the visual saliency. However, it further exploits the visual saliency map for weighting the spatial similarity pooling. Furthermore, [3] also explores the influence of different saliency models on the performance of the proposed image quality measure.

Adopting the advances in machine learning and data science, IQA methods following a third, purely *data driven* strategy have been proposed recently. So far, data driven approaches were mainly developed for the domain of NR IQA [14, 15, 16, 17], but they have also been adapted in the context of FR IQA [18].

The feature similarity index (FSIM) [2], proposed in 2011, has since then been one of the most successful and influential FR image quality metrics. It also shares certain conceptual similarities with the newly proposed HaarPSI measure. For the remainder of this section, we will hence examine FSIM in a little more detail. FSIM combines two feature maps derived from the phase congruency measure [19] and the local gradients of the reference and the distorted image, respectively, in order to estimate the perceived quality. For a grayscale image $f \in \ell^2(\mathbb{Z}^2)$, the gradient map is defined by

$$\mathrm{G_f}[x] = \sqrt{((g^{\mathrm{hor}} * f)[x])^2 + ((g^{\mathrm{ver}} * f)[x])^2}, \tag{1}$$

where $g^{\mathrm{hor}}$ and $g^{\mathrm{ver}}$ denote horizontal and vertical gradient filters (e.g. Sobel or Scharr filters), and $*$ denotes the two-dimensional convolution operator. The algorithm used by the authors of FSIM in order to compute the phase congruency map was developed by Peter Kovesi [20] and contains several non-trivial operations, such as adaptive soft thresholding. However, in its essence, the phase congruency map of a grayscale image $f$ can be described by

$$\mathrm{PC_f}[x] \approx \frac{|\sum_n (g_n^{\mathrm{c}} * f)[x]|}{\sum_n |(g_n^{\mathrm{c}} * f)[x]|}, \tag{2}$$

where $(g_n^{\mathrm{c}})_n$ is a set of differently scaled and oriented complex-valued wavelet filters. The idea behind (2) is that if the obtained complex-valued wavelet coefficients have the same phase at a location $x$, taking the absolute value of the sum is the same as taking the sum of the absolute values. In this case, $\mathrm{PC_f}[x]$ will be close to or precisely 1.

To assess local similarities between two images with respect to the maps defined in (1) and (2), FSIM - like many other image quality metrics - uses a simple comparison function for scalar values that already appeared in [1], namely

$$\mathrm{S}(a, b, C) = \frac{2ab + C}{a^2 + b^2 + C}, \tag{3}$$

where the constant $C > 0$ provides stability in the case that $a^2 + b^2$ is close to zero. Using (3), a local feature similarity map is defined for two grayscale images $f_1, f_2 \in \ell^2(\mathbb{Z}^2)$ by

$$\mathrm{FS_{f_1,f_2}}[x] = \mathrm{S}\left(\mathrm{G_{f_1}}[x], \mathrm{G_{f_2}}[x], C_1\right)^{\alpha} \cdot \mathrm{S}\left(\mathrm{PC_{f_1}}[x], \mathrm{PC_{f_2}}[x], C_2\right)^{\beta}, \tag{4}$$

with constants $C_1, C_2 > 0$ and exponents $\alpha, \beta > 0$. Based on the assumption that the human visual system is especially sensitive towards structures like edges and ridges, at which the phases of the Fourier components are in congruency (see e.g. [21]), the phase congruency map is not

only used in (4) but also applied to determine the relative importance of different image areas with respect to human perception. Eventually, the feature similarity index is computed by taking the weighted mean of all local feature similarities, where the phase congruency map is used as a weight function, that is

$$\mathrm{FSIM}_{\mathrm{f}_1,\mathrm{f}_2} = \frac{\sum_x \mathrm{FS}_{\mathrm{f}_1,\mathrm{f}_2}[x] \cdot \mathrm{PC}_{\mathrm{f}_1,\mathrm{f}_2}[x]}{\sum_x \mathrm{PC}_{\mathrm{f}_1,\mathrm{f}_2}[x]}, \tag{5}$$

where

$$\mathrm{PC}_{\mathrm{f}_1,\mathrm{f}_2}[x] = \max\left(\mathrm{PC}_{\mathrm{f}_1}[x], \mathrm{PC}_{\mathrm{f}_2}[x]\right). \tag{6}$$

The original publication of FSIM also proposes a generalization to color images defined in the YIQ color space, named FSIMC. In the YIQ space, the Y channel encodes luminance information, while the I and Q channels encode chromatic information. Color images defined in the RGB color space can easily be transformed to the YIQ space with a linear mapping, namely

$$\begin{bmatrix} f^{\mathrm{Y}} \\ f^{\mathrm{I}} \\ f^{\mathrm{Q}} \end{bmatrix} \approx \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \cdot \begin{bmatrix} f^{\mathrm{R}} \\ f^{\mathrm{G}} \\ f^{\mathrm{B}} \end{bmatrix}. \tag{7}$$

FSIMC simply incorporates the chroma channels I and Q into the local feature similarity measure (4). The gradient maps as well as the phase congruency maps are purely based on the luminance channel Y in FSIMC and FSIM alike.

We think that the mathematical structure of HaarPSI described in the upcoming Section 3 can best be understood in comparison with the structure of and the concepts behind FSIM. In particular, HaarPSI can be considered a significantly simplified version of FSIM that not only requires much less computational effort but also clearly outperforms FSIM on all benchmark databases in Section 4.

## 3 The Haar Wavelet-Based Perceptual Similarity Index

The basic idea of HaarPSI is to construct feature maps in the spirit of (1) as well as a weight function similar to (2) by considering a single wavelet filterbank. The response of any high-frequency wavelet filter will look similar to the response yielded by a classical gradient filter, like the Sobel operator. Furthermore, the phase congruency measure used as a weight function in FSIM is computed directly from the output of a multi-scale complex-valued wavelet filterbank, as illustrated in equation (2). This gives a strong intuition that it should indeed be possible to define a similarity measure based on a single set of wavelet filters, that at least matches the performance of FSIM on benchmark databases, but requires significantly less computational effort.

The wavelet we choose for this endeavor is the so-called Haar wavelet, which was already proposed in 1910 by Alfred Haar [22] and is arguably the simplest and computationally most efficient wavelet there is. The one-dimensional Haar filters are given by

$$h_1^{\mathrm{1D}} = \sqrt{2} \cdot [1, 1] \text{ and } g_1^{\mathrm{1D}} = \sqrt{2} \cdot [-1, 1], \tag{8}$$

where $h_1^{\mathrm{1D}}$ denotes the low-pass scaling filter and $g_1^{\mathrm{1D}}$ the corresponding high-pass wavelet filter. For any scale $j \in \mathbb{N}$, we can construct two-dimensional Haar filters by setting

$$g_j^{(1)} = g_j^{\mathrm{1D}} \otimes h_j^{\mathrm{1D}},$$
$$g_j^{(2)} = h_j^{\mathrm{1D}} \otimes g_j^{\mathrm{1D}},$$

where $\otimes$ denotes the outer product and the one-dimensional filters $h_j^{\mathrm{1D}}$ and $g_j^{\mathrm{1D}}$ are given for $j > 1$ by

$$g_j^{\mathrm{1D}} = h_1^{\mathrm{1D}} * (g_{j-1}^{\mathrm{1D}})_{\uparrow 2},$$
$$h_j^{\mathrm{1D}} = h_1^{\mathrm{1D}} * (h_{j-1}^{\mathrm{1D}})_{\uparrow 2},$$

where $\uparrow 2$ is the dyadic upsampling operator, and $*$ denotes the one-dimensional convolution operator. Note that $g_j^{(1)}$ responds to horizontal structures, while $g_j^{(2)}$ picks up vertical structures. All eight Haar filters used in the computation of HaarPSI are shown in Figure 1.

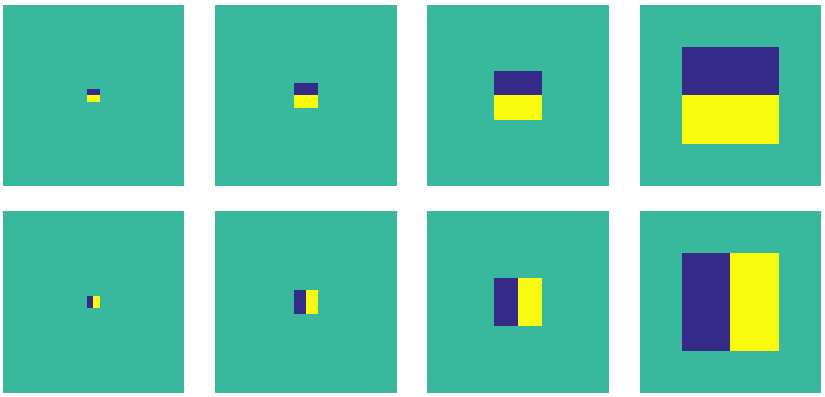


Figure 1: The eight Haar wavelet filters applied in the Haar wavelet-based perceptual similarity index (11).

For two grayscale images $f_1, f_2 \in \ell^2(\mathbb{Z}^2)$, we define a local similarity measure based on the first two stages of a two-dimensional discrete Haar wavelet transform, namely

$$\mathrm{HS}_{\mathrm{f}_1,\mathrm{f}_2}^{(\mathrm{k})}[x] = \prod_{j=1}^{2} \mathrm{S}\left(\left|(g_j^{(\mathrm{k})} * f_1)[x]\right|, \left|(g_j^{(\mathrm{k})} * f_2)[x]\right|, C_1\right), \tag{9}$$

where $C_1 > 0$ is a constant, $k \in \{1, 2\}$ selects either horizontal or vertical filters, S denotes the scalar similarity measure (3), and $*$ is the two-dimensional convolution operator. The local similarity measure (9) can be seen as an analog to (4). However, the Haar-based measure refrains from including a feature map like the phase congruency map (2), that is based on various partly nonlinear computations, involving a total of 16 complex-valued wavelet filters. Instead, (9) only relies on two scales of a real-valued Haar wavelet transform and hence requires significantly less computational effort and is easier to be included in any optimization process. A visualization of a local similarity map derived from (9) can be found in the second row of Figure 2.

Furthermore, a simple weight function can be defined by taking the sum of wavelet coefficients across four scales of a discrete Haar transform:

$$\mathrm{W}_{\mathrm{f}}^{(\mathrm{k})}[x] = \left|\sum_{j=1}^{4} (g_j^{(\mathrm{k})} * f)[x]\right|, \tag{10}$$

where $k \in \{1, 2\}$ again differentiates between horizontal and vertical filters. Note that this weight function is essentially the quotient (2) reduced to its numerator. Figure 2 shows an example of weights computed from a natural image by (10).

Using the weight function (10), the Haar similarity index for two grayscale images $f_1, f_2$ is finally given as the weighted average of the local similarity measure (9), that is,

$$\mathrm{HaarPSI}_{\mathrm{f}_1,\mathrm{f}_2} = \left( \frac{\sum_x \sum_{k=1}^{2} \left( \mathrm{HS}^{(\mathrm{k})}_{\mathrm{f}_1,\mathrm{f}_2}[x] \right)^{\alpha} \cdot \mathrm{W}^{(\mathrm{k})}_{\mathrm{f}_1,\mathrm{f}_2}[x]}{\sum_x \sum_{k=1}^{2} \mathrm{W}^{(\mathrm{k})}_{\mathrm{f}_1,\mathrm{f}_2}[x]} \right)^{\frac{1}{\alpha}}, \tag{11}$$

with an exponent $\alpha > 0$ and

$$\mathrm{W}^{(\mathrm{k})}_{\mathrm{f}_1,\mathrm{f}_2}[x] = \max(\mathrm{W}^{(\mathrm{k})}_{\mathrm{f}_1}[x], \mathrm{W}^{(\mathrm{k})}_{\mathrm{f}_2}[x]). \tag{12}$$

Please note that due to the monotonicity of the power function in the interval $[0, 1]$, omitting the exponent $\frac{1}{\alpha}$ in (11) would have no effect on the rank order-based correlations with human opinion scores reported in Section 4.

Analogous to FSIM, HaarPSI can be extended to color images in the YIQ color space by including the chroma channels I and Q in the local similarity measure (9). Formally, this generalization is given by

$$\mathrm{HaarPSIC}_{\mathrm{f}_1,\mathrm{f}_2} = \left( \frac{\sum_x \sum_{k=1}^{2} \left( \mathrm{HSC}^{(\mathrm{k})}_{\mathrm{f}_1,\mathrm{f}_2}[x] \right)^{\alpha} \cdot \mathrm{W}^{(\mathrm{k})}_{\mathrm{f}_1^{\mathrm{Y}},\mathrm{f}_2^{\mathrm{Y}}}[x]}{\sum_x \sum_{k=1}^{2} \mathrm{W}^{(\mathrm{k})}_{\mathrm{f}_1^{\mathrm{Y}},\mathrm{f}_2^{\mathrm{Y}}}[x]} \right)^{\frac{1}{\alpha}}, \tag{13}$$

with a single exponent $\alpha > 0$ and a chroma-sensitive local similarity measure

$$\begin{aligned} \mathrm{HSC}^{(\mathrm{k})}_{\mathrm{f}_1,\mathrm{f}_2}[x] = \prod_{j=1}^{2} \mathrm{S}\left( \left| (g_j^{(\mathrm{k})} * f_1^{\mathrm{Y}})[x] \right|, \left| (g_j^{(\mathrm{k})} * f_2^{\mathrm{Y}})[x] \right|, C_1 \right) \\ \cdot \mathrm{S}((m * f_1^{\mathrm{I}})[x], (m * f_2^{\mathrm{I}})[x], C_2) \\ \cdot \mathrm{S}((m * f_1^{\mathrm{Q}})[x], (m * f_2^{\mathrm{Q}})[x], C_2), \end{aligned} \tag{14}$$

with constants $C_1, C_2 > 0$ and a $2 \times 2$ mean filter $m$.

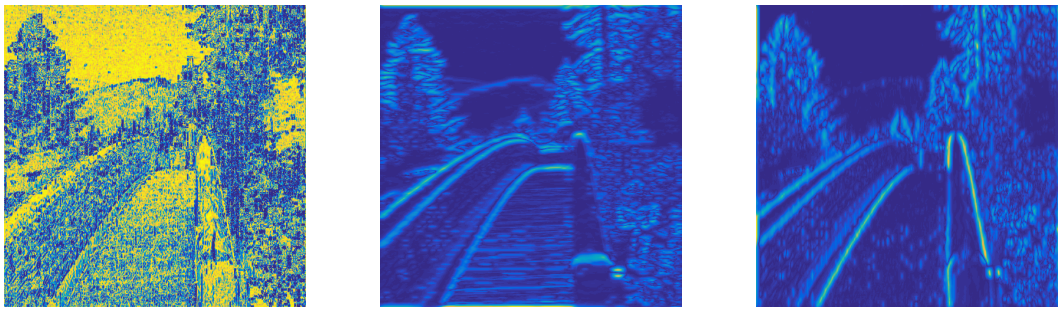
The grayscale version of HaarPSI only requires two parameters to be selected, namely $C_1$ in (9) and $\alpha$ in (11), while HaarPSIC uses a second constant $C_2$ in (14). These parameters were optimized by the authors to yield a superior overall performance on all four databases considered in section 4 and chosen to be $C_1 = 40$, $\alpha = 0.03$ and $C_2 = 250$. However, it should be noted that for other databases or specific applications, different values might still be favorable (see Figure 4).

# 4 Experimental Procedure and Results

The consistency of HaarPSI with the human perception of image quality was evaluated and compared with most of the image quality metrics discussed in Section 2 with four large publicly available benchmark databases of quality-annotated images. Those databases differ in the number of reference images, the number of distortion magnitudes and types, the number of observers, the level of control of the viewing conditions, and the stimulus presentation procedure.

The LIVE database [23] contains 29 reference color images and 779 distorted images that were perturbed by JPEG compression, JPEG 2000 compression, additive Gaussian white noise,

(a) reference $f_1$ (b) distorted $f_2$ (c) $\mathrm{HS}^{(1)}_{\mathrm{f}_1,\mathrm{f}_2}$



(d) $\mathrm{HS}^{(2)}_{\mathrm{f}_1,\mathrm{f}_2}$ (e) $\mathrm{W}^{(1)}_{\mathrm{f}_1,\mathrm{f}_2}$ (f) $\mathrm{W}^{(2)}_{\mathrm{f}_1,\mathrm{f}_2}$

Figure 2: (a) An undistorted reference image. (b) The reference image distorted by the JPEG compression algorithm. (c) The horizontal local similarity map (11) ($k = 1$). (d) The vertical local similarity map (11) ($k = 2$). (e) The horizontal weight function (10) ($k = 1$). (f) The vertical weight function (10) ($k = 2$). Yellow indicates a high similarity in (c) and (d) and a high relative importance of an area in (e) and (f), while dark blue represents the respective opposite. The images (a) and (b) were taken from the CSIQ database [8].

Table 1: Spearman Rank Order Correlations of IQA Metrics With Human Mean Opinion Scores

| | PSNR | VIF | SSIM | MS-SSIM | GSM | MAD | SR-SIM | FSIM | VSI | HaarPSI |
|---|---|---|---|---|---|---|---|---|---|---|
| Grayscale Images | | | | | | | | | | |
| LIVE | 0.8756 | 0.9636 | 0.9479 | 0.9513 | 0.9561 | 0.9672 | 0.9619 | 0.9634 | 0.9534 | **0.9675** |
| TID 2008 | 0.5531 | 0.7491 | 0.7749 | 0.8542 | 0.8504 | 0.8340 | 0.8913 | 0.8804 | 0.8830 | **0.9042** |
| TID 2013 | 0.6394 | 0.6769 | 0.7417 | 0.7859 | 0.7946 | 0.7807 | 0.8075 | 0.8022 | 0.8048 | **0.8129** |
| CSIQ | 0.8058 | 0.9195 | 0.8756 | 0.9133 | 0.9108 | 0.9466 | 0.9319 | 0.9242 | 0.9372 | **0.9525** |
| Color Images | | | | | | | | | | |
| | PSNR | VIF | SSIM | MS-SSIM | GSM | MAD | SR-SIM | FSIM | VSI | HaarPSI |
| LIVE | 0.8756 | 0.9636 | 0.9479 | 0.9513 | 0.9561 | 0.9672 | 0.9619 | 0.9645 | 0.9524 | **0.9685** |
| TID 2008 | 0.5531 | 0.7491 | 0.7749 | 0.8542 | 0.8504 | 0.8340 | 0.8913 | 0.8840 | 0.8979 | **0.9079** |
| TID 2013 | 0.6394 | 0.6769 | 0.7417 | 0.7859 | 0.7946 | 0.7807 | 0.8075 | 0.8510 | **0.8965** | 0.8791 |
| CSIQ | 0.8058 | 0.9195 | 0.8756 | 0.9133 | 0.9108 | 0.9466 | 0.9319 | 0.9310 | 0.9423 | **0.9567** |

The highest correlation in each row is written in boldface.

Gaussian blurring as well as JPEG 2000 compressed images that have been transmitted over a simulated Rayleigh fading channel. Each distortion is introduced at five to six different levels of magnitude. On average, about 23 subjects evaluated the quality of each image with respect to the reference image. The viewing conditions were fairly controlled for in terms of viewing distance. Ratings were collected in a double stimulus manner.

The TID 2008 database [24] comprises 25 colored reference images and 1700 degraded images, that had been subject to a wide range of distortions, including various types of noise, blur,

JPEG and JPEG 2000 compression, transmission errors, local image distortions, as well as luminance and contrast changes. Subjective ratings were gathered by comparisons. The results from several viewing conditions of experiments in three different labs and on the internet were averaged. TID 2008 was later extended to TID 2013 [25], which added new types of distortions, which are mostly of a chromatic nature. In total, TID 2013 contains 3000 differently distorted images.

The CSIQ database [8] is based on 30 reference color images. Six different types of distortions (JPEG compression, JPEG 2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring) at four to five different degradation magnitudes were applied to the reference images. The viewing distance was controlled. Images were presented on a monitor array and subjects were asked to place all distorted versions of one reference image according to its perceived quality.

For all databases, we use the Spearman rank oder correlation coefficient (SROCC) to measure the consistency of the similarity scores computed by an image quality metric with the mean opinion scores reported by human participants. The first step in computing the SROCC is to construct two sequences of integers by mapping each distorted image from a database to the rank respectively induced by the image quality metric and the human opinion scores. The SROCC is then defined as the Pearson correlation coefficient of these two sequences. If a numerical similarity measure is perfectly consistent with the subjective human opinion, the ranks will be the same for each distorted image and the SROCC will be exactly 1.

All four databases only contain color images. However, out of the metrics considered in our experiments, only FSIM and HaarPSI are defined for both grayscale and color images, while the visual saliency-based index (VSI) was specifically designed for color images. All other similarity measures considered in our experiments only accept grayscale images as inputs. To reflect these differing designs, all methods were tested on all databases once with the original color images and once with grayscale conversions. The correlation coefficients of all ten considered similarity measures with the human mean opinion scores for the LIVE image database, TID 2008, TID 2013 and the CSIQ database are compiled in Table 1.

Table 2 provides a quick impression of the overall performance of each metric. It depicts the average SROCC of each metric with respect to all four databases as well as the mean execution time in milliseconds. The average execution time was measured on a Intel Core i7-4790 CPU clocked at 3.60 GHz. Each quality measure was computed ten times for ten different pairs of randomly generated $512 \times 512$ pixel images.

Table 2: Mean SROCC and Execution Time

| | Color Images | | Grayscale Images | |
|---|---|---|---|---|
| | SROCC | Time (ms) | SROCC | Time (ms) |
| HaarPSI | 0.9280 | 30 | 0.9093 | 16 |
| VSI | 0.9223 | 75 | 0.8946 | 75 |
| FSIM | 0.9076 | 135 | 0.8925 | 116 |
| SR-SIM | 0.8982 | 10 | 0.8982 | 10 |
| MAD | 0.8821 | 880 | 0.8821 | 855 |
| GSM | 0.8780 | 8 | 0.8780 | 7 |
| MS-SSIM | 0.8762 | 29 | 0.8762 | 24 |
| SSIM | 0.8350 | 7 | 0.8350 | 5 |
| VIF | 0.8273 | 407 | 0.8273 | 394 |
| PSNR | 0.7185 | 2 | 0.7185 | 1 |

A high correlation with the mean opinion scores annotated to the distorted images of a large database containing many different types and degrees of distortions is arguably the best

indicator of an image quality measure's consistency with human perception. However, for certain applications like compression or denoising, it could be more important to know if an image quality metric has a high correlation the human experience *within* a single distortion class. Table 3 depicts the SROC coefficients for all image quality metrics, when only subsets of databases containing specific distortions like Gaussian blur or JPEG transmission errors are considered.

Table 3: Spearman Rank Order Correlations of IQA Metrics With Human Mean Opinion Scores

| | | Color Images | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR | VIF | SSIM | MS-SSIM | GSM | MAD | SR-SIM | FSIM | VSI | HaarPSI |
| LIVE | jpg2k | 0.8954 | 0.9696 | 0.9614 | 0.9627 | 0.9700 | 0.9692 | 0.9700 | **0.9724** | 0.9604 | 0.9680 |
| | jpg | 0.8809 | **0.9846** | 0.9764 | 0.9815 | 0.9778 | 0.9786 | 0.9823 | 0.9840 | 0.9761 | 0.9822 |
| | gwn | 0.9854 | 0.9858 | 0.9694 | 0.9733 | 0.9774 | **0.9873** | 0.9812 | 0.9716 | 0.9835 | 0.9863 |
| | gblur | 0.7823 | **0.9728** | 0.9517 | 0.9542 | 0.9518 | 0.9510 | 0.9660 | 0.9708 | 0.9527 | 0.9613 |
| | ff | 0.8907 | **0.9650** | 0.9556 | 0.9471 | 0.9402 | 0.9589 | 0.9466 | 0.9519 | 0.9430 | 0.9555 |
| TID 2008 | gwn | 0.9070 | 0.8797 | 0.8107 | 0.8086 | 0.8606 | 0.8386 | 0.8989 | 0.8758 | **0.9229** | 0.9162 |
| | gwnc | 0.8995 | 0.8757 | 0.8029 | 0.8054 | 0.8091 | 0.8255 | 0.8957 | 0.8931 | **0.9118** | 0.9002 |
| | scn | 0.9170 | 0.8698 | 0.8145 | 0.8209 | 0.8941 | 0.8678 | 0.9084 | 0.8711 | 0.9296 | **0.9357** |
| | mn | 0.8515 | **0.8683** | 0.7795 | 0.8107 | 0.7452 | 0.7336 | 0.7881 | 0.8264 | 0.7734 | 0.7956 |
| | hfn | **0.9270** | 0.9075 | 0.8729 | 0.8694 | 0.8945 | 0.8864 | 0.9195 | 0.9156 | 0.9253 | 0.9170 |
| | in | **0.8724** | 0.8327 | 0.6732 | 0.6907 | 0.7235 | 0.0650 | 0.7678 | 0.7719 | 0.8298 | 0.8113 |
| | qn | 0.8696 | 0.7970 | 0.8531 | 0.8589 | 0.8800 | 0.8160 | 0.8348 | 0.8726 | 0.8731 | **0.8831** |
| | gblr | 0.8697 | 0.9540 | 0.9544 | 0.9563 | **0.9600** | 0.9196 | 0.9551 | 0.9472 | 0.9529 | 0.8878 |
| | den | 0.9416 | 0.9161 | 0.9530 | 0.9582 | **0.9725** | 0.9433 | 0.9666 | 0.9618 | 0.9693 | 0.9714 |
| | jpg | 0.8717 | 0.9168 | 0.9252 | 0.9322 | 0.9393 | 0.9275 | 0.9393 | 0.9294 | **0.9616** | 0.9555 |
| | jpg2k | 0.8132 | 0.9709 | 0.9625 | 0.9700 | 0.9758 | 0.9707 | 0.9809 | 0.9780 | 0.9848 | **0.9864** |
| | jpgt | 0.7516 | 0.8585 | 0.8678 | 0.8681 | 0.8790 | 0.8661 | 0.8881 | 0.8756 | **0.9160** | 0.8930 |
| | jpg2kt | 0.8309 | 0.8501 | 0.8577 | 0.8606 | 0.8936 | 0.8394 | 0.8902 | 0.8555 | 0.8942 | **0.9061** |
| | pn | 0.5815 | 0.7619 | 0.7107 | 0.7377 | 0.7386 | **0.8287** | 0.7659 | 0.7514 | 0.7699 | 0.8112 |
| | bdist | 0.6193 | 0.8324 | 0.8462 | 0.7546 | **0.8862** | 0.7970 | 0.7798 | 0.8464 | 0.6295 | 0.8064 |
| | ms | 0.6957 | 0.5096 | 0.7231 | **0.7338** | 0.7190 | 0.5163 | 0.5704 | 0.6554 | 0.6714 | 0.6063 |
| | ctrst | 0.5859 | **0.8188** | 0.5246 | 0.6381 | 0.6691 | 0.2723 | 0.6475 | 0.6510 | 0.6557 | 0.6415 |
| TID 2013 | gwn | 0.9291 | 0.8994 | 0.8671 | 0.8646 | 0.9064 | 0.8843 | 0.9251 | 0.9101 | **0.9460** | 0.9365 |
| | gwnc | **0.8981** | 0.8299 | 0.7726 | 0.7730 | 0.8175 | 0.8019 | 0.8562 | 0.8537 | 0.8705 | 0.8579 |
| | scn | 0.9200 | 0.8835 | 0.8515 | 0.8544 | 0.9158 | 0.8911 | 0.9223 | 0.8900 | 0.9367 | **0.9396** |
| | mn | 0.8323 | **0.8450** | 0.7767 | 0.8073 | 0.7293 | 0.7380 | 0.7855 | 0.8094 | 0.7697 | 0.7853 |
| | hfn | 0.9140 | 0.8972 | 0.8634 | 0.8604 | 0.8869 | 0.8876 | 0.9131 | 0.9040 | **0.9200** | 0.9102 |
| | in | **0.8968** | 0.8537 | 0.7503 | 0.7629 | 0.7965 | 0.2769 | 0.8280 | 0.8251 | 0.8741 | 0.8559 |
| | qn | 0.8808 | 0.7854 | 0.8657 | 0.8706 | 0.8841 | 0.8514 | 0.8497 | 0.8807 | 0.8748 | **0.8872** |
| | gblr | 0.9149 | 0.9650 | 0.9668 | 0.9673 | **0.9689** | 0.9319 | 0.9622 | 0.9551 | 0.9612 | 0.9120 |
| | den | 0.9480 | 0.8911 | 0.9254 | 0.9268 | 0.9432 | 0.9252 | 0.9398 | 0.9330 | **0.9484** | 0.9460 |
| | jpg | 0.9189 | 0.9192 | 0.9200 | 0.9265 | 0.9284 | 0.9217 | 0.9396 | 0.9339 | 0.9541 | **0.9556** |
| | jpg2k | 0.8840 | 0.9516 | 0.9468 | 0.9504 | 0.9602 | 0.9511 | 0.9672 | 0.9589 | **0.9706** | 0.9703 |
| | jpgt | 0.7685 | 0.8409 | 0.8493 | 0.8475 | 0.8512 | 0.8283 | 0.8543 | 0.8610 | **0.9216** | 0.8886 |
| | jpg2kt | 0.8883 | 0.8761 | 0.8828 | 0.8889 | 0.9182 | 0.8788 | 0.9165 | 0.8919 | 0.9228 | **0.9239** |
| | pn | 0.6863 | 0.7720 | 0.7821 | 0.7968 | 0.8130 | **0.8315** | 0.7967 | 0.7937 | 0.8060 | 0.8227 |
| | bdist | 0.1552 | 0.5306 | 0.5720 | 0.4801 | **0.6418** | 0.2812 | 0.4722 | 0.5532 | 0.1713 | 0.4598 |
| | ms | 0.7671 | 0.6276 | 0.7752 | **0.7906** | 0.7875 | 0.6450 | 0.6562 | 0.7487 | 0.7700 | 0.7156 |
| | ctrst | 0.4400 | **0.8386** | 0.3775 | 0.4634 | 0.4857 | 0.1972 | 0.4696 | 0.4679 | 0.4754 | 0.4569 |
| | ccs | 0.0766 | 0.3099 | 0.4141 | 0.4099 | 0.3578 | 0.0575 | 0.3117 | **0.8359** | 0.8100 | 0.6307 |
| | mgn | 0.8905 | 0.8468 | 0.7803 | 0.7786 | 0.8348 | 0.8409 | 0.8781 | 0.8569 | **0.9117** | 0.8892 |
| | cn | 0.8411 | 0.8946 | 0.8566 | 0.8528 | 0.9124 | 0.9064 | **0.9259** | 0.9135 | 0.9243 | 0.9224 |
| | lcni | 0.9145 | 0.9204 | 0.9057 | 0.9068 | 0.9563 | 0.9443 | **0.9608** | 0.9485 | 0.9564 | 0.9559 |
| | icqd | **0.9269** | 0.8414 | 0.8542 | 0.8555 | 0.8973 | 0.8745 | 0.8810 | 0.8815 | 0.8839 | 0.9026 |
| | cha | 0.8872 | 0.8848 | 0.8775 | 0.8784 | 0.8823 | 0.8310 | 0.8758 | **0.8925** | 0.8906 | 0.8709 |
| | ssr | 0.9042 | 0.9353 | 0.9461 | 0.9483 | **0.9668** | 0.9567 | 0.9613 | 0.9576 | 0.9628 | 0.9642 |
| CSIQ | gwn | 0.9363 | 0.9575 | 0.8974 | 0.9471 | 0.9440 | 0.9541 | 0.9628 | 0.9359 | 0.9636 | **0.9659** |
| | jpeg | 0.8881 | **0.9705** | 0.9546 | 0.9634 | 0.9632 | 0.9615 | 0.9671 | 0.9664 | 0.9618 | 0.9684 |
| | jpg2k | 0.9362 | 0.9672 | 0.9606 | 0.9683 | 0.9648 | 0.9752 | 0.9773 | 0.9704 | 0.9694 | **0.9794** |
| | gpn | 0.9339 | 0.9511 | 0.8922 | 0.9331 | 0.9387 | 0.9570 | 0.9520 | 0.9370 | **0.9638** | 0.9607 |
| | gblr | 0.9291 | 0.9745 | 0.9609 | 0.9711 | 0.9589 | 0.9682 | 0.9767 | 0.9729 | 0.9679 | **0.9775** |
| | ctrst | 0.8621 | 0.9345 | 0.7922 | 0.9526 | 0.9354 | 0.9207 | **0.9528** | 0.9438 | 0.9504 | 0.9500 |

The highest correlation in each row is written in boldface.

Finally, it should be noted that for all results reported in this section, HaarPSI, as well as most other image quality metrics, was preprocessing each image by convolving it with a $2 \times 2$ mean filter as well as a subsequent dyadic subsampling step, to simulate the viewing distance between the participants and the presented images in the experimental setup.

# 5 Discussion

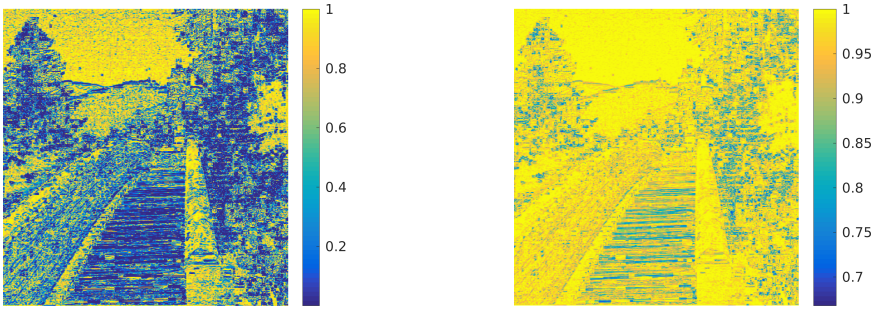
HaarPSI is a novel and computationally inexpensive image quality measure based solely on the coefficients of four stages of a discrete Haar wavelet transform. Its validity with respect to the human perception of image quality was tested on four large databases containing more than 5000 differently distorted images, with very promising results. When restricted to grayscale conversions, HaarPSI outperforms all state-of-the-art algorithms considered in Section 4 on all four databases, while for color images, it only comes second to VSI when tested on TID 2013 (see Table 1). Along with its simple computational structure and its comparatively short execution time, this suggests a high applicability of HaarPSI in real world optimization tasks. In particular, image quality metrics like PSNR, SSIM, or SR-SIM, that outperform HaarPSI with respect to speed, achieve considerably inferior correlations with human opinion scores (see Table 2). Regarding the applicability of HaarPSI in specific optimization tasks, we would like to mention that HaarPSI has consistently high correlations with human opinion scores throughout all databases with respect to distortions caused by the JPEG and JPEG 2000 compression algorithms (see Table 3).

As it was already noted in Section 3, HaarPSI can conceptually be understood as a simplified version of FSIM. Both HaarPSI and FSIM rely on the construction of two maps, where one map measures local similarities between a reference image and a distorted image and the other map assesses the relative importance of image areas. However, in HaarPSI, both maps are defined only in terms of a single Haar wavelet filterbank, while FSIM utilizes an implementation of the phase congruency measure that not only requires the images to be convolved with 16 complex-valued filters but also contains several non-trivial computational steps, like adaptive thresholding. Furthermore, FSIM uses the phase congruency measure both as a weight function in (5) and as a part of the local similarity measure (4). In HaarPSI, the weight function (10) and the local similarity measure (9) are strictly separated. These conceptual simplifications not only decrease the execution time by a factor of approximately five (see Table 2), but also pave the way for significantly higher correlations with opinion scores of human viewers (see Tables 1 and 3). In particular, it seems to be an important step to split the measurements regarding horizontal structures and vertical structures in both the local similarity measure (9) and the weight function (10).

Other visual quality metrics emphasize the application of concepts motivated by neurophysiological findings, like phase congruency or visual saliency. Besides the fact that simple cells in the primary visual cortex were found to serve as edge detectors and hence yield responses structurally not too different from the responses of Haar wavelet filters, HaarPSI cannot be said to have a strong connection to computational models of human perception. That being said, it is interesting to note that the local similarity measure (9) only considers the high-frequency responses of the Haar wavelet transform, while the weight function (10) strongly depends on the low-frequency information yielded by the third and fourth scale of the discrete Haar wavelet transform. This could indicate that the human visual system recognizes changes more significantly in the high-frequency range but uses low-frequency information to judge the overall importance of an image area.

The definition of HaarPSI (11) contains two free parameters, namely a constant $C_1$ that stabilizes the local similarity measure (9) when the denominator in (3) is close to 0 and an

exponent $\alpha$. HaarPSIC adds a third constant, $C_2$, that is required for the inclusion of the chroma channels in the color-sensitive local similarity measure (14). These parameters were fixed by the authors at $C_1 = 40$, $C_2 = 250$ and $\alpha = 0.03$, in order to maximize the correlations with human mean opinion scores with respect to all four databases considered in Section 4. While the choices for the constants $C_1$ and $C_2$ are comparable to the values given to similar parameters in other metrics like FSIM or VSI, it was quite surprising to see that an optimal choice for the exponent $\alpha$ in (11) and (13) should be as small as 0.03. By selecting an exponent that close to zero, all values in the local similarity map (9) are being mapped to a value rather close to one. In particular, the ascent towards one is much steeper for values near zero. Thus, with $\alpha = 0.03$, only locations with a strong dissimilarity between the reference and the distorted image will play a significant role in the overall measures (11) and (13) (see Figure 3). This slightly resembles the effects of a discontinuous step-function or of a logistic function, both of which are frequently used as transfer functions in artificial neural networks. It is quite possible that a more exhaustive examination of this connection could lead to even better results for HaarPSI and similar metrics. For instance, it might be beneficial to actually use a step function or a logistic function instead of the exponent $\alpha$ in (11) and (13), or to directly apply such functions to the coefficients obtained from the discrete wavelet transform.

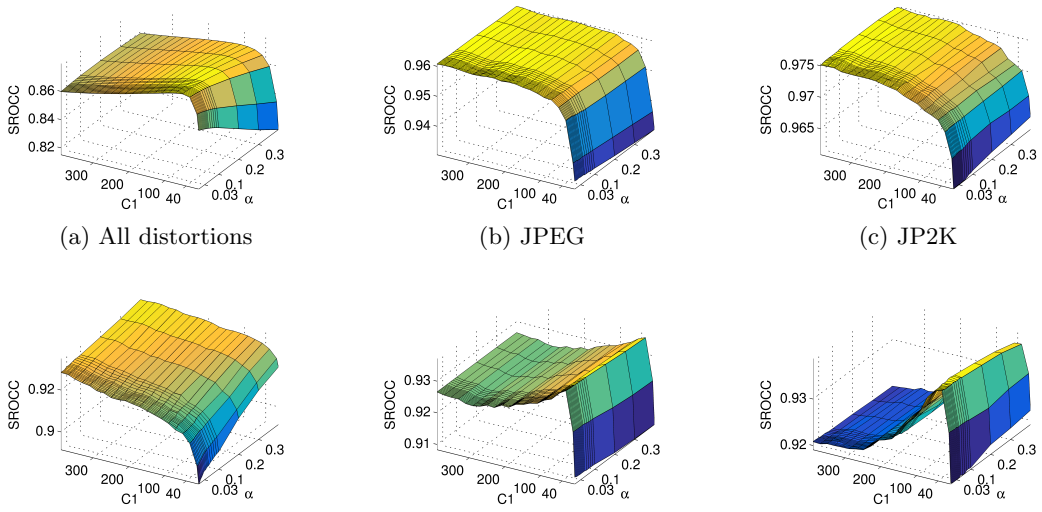



(a) $\mathrm{HS}^{(1)}_{\mathrm{f}_1,\mathrm{f}_2}$ (b) $\left(\mathrm{HS}^{(1)}_{\mathrm{f}_1,\mathrm{f}_2}\right)^{\alpha}$

Figure 3: (a) The local similarity measure (9) already shown in Figure 2. (b) Effect of the exponent $\alpha = 0.03$ when applied entry-wise to the local similarity map shown on the left.

As can be seen in Table 3, HaarPSI is sometimes outperformed by other IQA methods, when only specific types of distortions are being considered. However, even in these cases, correlations comparable or superior to the correlations of other state-of-the-art similarity measures might be achieved by tuning the constants $C_1$, $C_2$, and $\alpha$, which have originally been selected to optimize the overall performance of HaarPSI. The influence of $C_1$ and $\alpha$ on the correlation with human opinion scores in the case of TID 2013 is shown in Figure 4 with respect to the overall performance as well as for five specific distortions (JPEG, JP2K, Gaussian blur, additive white Gaussian noise, and spatially correlated noise). Note that $C_2 = 250$ remained constant throughout all experiments depicted in Figure 4. Figure 4a reinforces the claim that setting $C_1 = 40$ and $\alpha = 0.03$ optimizes HaarPSI with respect to distortion agnostic IQA. Figure 4b indicates that increasing $C_1$ also increases the SROCC for JPEG-distorted images, whereas the performance is rather independent of $\alpha$. IQA of images affected by JP2K or Gaussian Blur would benefit from increasing both $C_1$ and $\alpha$, as can be seen in Figures 4c and 4d. However, as shown in Figures 4a, 4e and 4f, increasing $C_1$ and $\alpha$ is not beneficial in general. For quality assessments of images degraded by Gaussian white noise or spatially correlated noise, increasing $C_1$ results in a sharp drop in the correlation with human opinion scores.

From a computational point of view, it is very beneficial to apply the Haar wavelet in a




(a) All distortions (b) JPEG (c) JP2K

(d) Blur (e) AWGN (f) SCN

Figure 4: Spearman rank order correlations in dependance of the parameters $C_1$ and $\alpha$ for images affected by (a) all distortions, (b) JPEG compression, (c) JP2K compression, (d) Gaussian Blur, (e) additive Gaussien white noise, and (f) spatially correlated noise. $C_2 = 250$ for all calculations. All correlations are with respect to TID2013.

method like HaarPSI instead of other wavelet filters. However, by simply changing $h_1^{1D}$ and $g_1^{1D}$ in (8) to the respective filters, the measure given in (11) can also be defined for other wavelets. Table 4 shows the performance of such measures based on selected Daubechies wavelets [26], symlets [27], coiflets [28] and the Cohen-Daubechies-Feauveau wavelet [29] with respect to the four databases considered in Section 4. It is interesting to see that Haar filters not only seem to be the computationally most efficient but also the qualitatively best choice for the measure (11).

Table 4: SROCC With Human Mean Opinion Scores For Different Wavelet Filters

| | Grayscale Images | | | | | |
|---|---|---|---|---|---|---|
| | Daub2PSI | Daub4PSI | Sym4PSI | CDFPSI | Coif1PSI | HaarPSI |
| LIVE | 0.9593 | 0.9507 | 0.9527 | 0.9507 | 0.9579 | **0.9675** |
| TID 2008 | 0.8852 | 0.8539 | 0.8746 | 0.8742 | 0.8828 | **0.9042** |
| TID 2013 | 0.8022 | 0.7819 | 0.7962 | 0.7946 | 0.8019 | **0.8129** |
| CSIQ | 0.9492 | 0.9453 | 0.9453 | 0.9413 | 0.9482 | **0.9525** |
| | Color Images | | | | | |
| | Daub2PSI | Daub4PSI | Sym4PSI | CDFPSI | Coif1PSI | HaarPSI |
| LIVE | 0.9623 | 0.9572 | 0.9601 | 0.9594 | 0.9618 | **0.9685** |
| TID 2008 | 0.8896 | 0.8612 | 0.8803 | 0.8822 | 0.8864 | **0.9079** |
| TID 2013 | 0.8700 | 0.8569 | 0.8675 | 0.8672 | 0.8719 | **0.8791** |
| CSIQ | **0.9580** | 0.9548 | 0.9561 | 0.9564 | 0.9561 | 0.9567 |

The highest correlation in each row is written in boldface.

A MATLAB function implementing HaarPSI will be available at `www.math.uni-bremen.de/cda/software.html`.

## References


[1] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Proc.*, 13(4):600–612, 2004.

[2] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Trans. Image Proc.*, 20(8):2378–2386, 2011.

[3] L. Zhang, Y. Shen, and H. Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, Oct 2014.

[4] A.B. Watson, R. Borthwick, and M. Taylor. Image quality and entropy masking. In *SPIE Proceedings*, volume 3016, pages 1–11, 1997.

[5] Scott J Daly. Application of a noise-adaptive contrast sensitivity function to image data compression. *Optical Engineering*, 29(8):977–987, 1990.

[6] Jeffrey Lubin. A human vision system model for objective picture quality measurements. *International Broadcasting Convention*, pages 498–503, 1997.

[7] Yuting Jia, Weisi Lin, and Ashraf A Kassim. Estimating just-noticeable distortion for video. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(7):820–829, 2006.

[8] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006–1–011006–21, 2010.

[9] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15:430–444, 2006.

[10] A. Liu, W. Lin, and M. Narwaria. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 21(4):1500–1512, April 2012.

[11] L. Zhang and H. Li. Sr-sim: A fast and high performance iqa index based on spectral residual. In *2012 19th IEEE International Conference on Image Processing*, pages 1473–1476, Sept 2012.

[12] Richard Dosselmann and Xue Dong Yang. A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, 5(1):81–91, 2011.

[13] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik. Multi-scale structural similarity fror image quality assessment. In *Proceedings of 37th IEEE Asilomar Conference on Signals, Systems and Computers*, 2003.

[14] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional Neural Networks for No-Reference Image Quality Assessment. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1733–1740, 2014.

[15] P. Ye and D. Doermann. No-Reference Image Quality Assessment Using Visual Codebooks. *IEEE Transactions on Image Processing*, 21(7):3129–3138, 2012.

[16] P. Zhang, W. Zhou, L. Wu, and H. Li. SOM: Semantic obviousness metric for image quality assessment. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2394–2402, 2015.

[17] S. Bosse, D. Maniry, T. Wiegand, and W. Samek. A deep neural network for image quality assessment. In *Image Processing (ICIP), 2016 IEEE International Conference on*, 2016.

[18] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek. Full-reference image quality assessment using neural networks. In *Int. Work. Qual. Multimed. Exp.*, 2016.

[19] Peter Kovesi. Phase congruency: A low-level image invariant. *Psychological Research*, 64:136–148, 2000.

[20] Peter D. Kovesi. Matlab and octave functions for computer vision and image processing. Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia. Available from `http://www.csse.uwa.edu.au/~pk/research/matlabfns/`.

[21] M. C. Morrone, J. R. Ross, D. C. Burr, and R. A. Owens. Mach bands are phase dependent. *Nature*, 324(6094):250–253, 1986.

[22] Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910.

[23] Hamid Rahim Sheikh, Zhou Wang, Lawrance Cormack, and Alan C. Bovik. Live image quality assessment database release 2. Available from `http://live.ece.utexas.edu/research/quality`.

[24] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10:30–45, 2009.

[25] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57 – 77, 2015.

[26] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996, 1988.

[27] Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

[28] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets ii: Variations on a theme. *SIAM J. Math. Anal.*, 24(2):499–519, 1993.

[29] A. Cohen, Ingrid Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5):485–560, 1992.