# Most Apparent Distortion: A Dual Strategy for Full-Reference Image Quality Assessment

Eric C. Larson and Damon M. Chandler

Image Coding and Analysis Lab, School of Electrical and Computer Engineering,
Oklahoma State University, Stillwater, OK 74078

## ABSTRACT

The mainstream approach to image quality assessment has centered around accurately modeling the *single* most relevant strategy employed by the human visual system (HVS) when judging image quality (e.g., detecting visible differences; extracting image structure/information). In this paper, we suggest that a single strategy may not be sufficient; rather, we advocate that the HVS uses *multiple* strategies to determine image quality. For images containing near-threshold distortions, the image is most apparent, and thus the HVS attempts to look *past* the image and look *for* the distortions (a detection-based strategy). For images containing clearly visible distortions, the distortions are most apparent, and thus the HVS attempts to look past the distortion and look for the image's subject matter (an appearance-based strategy). Here, we present a quality assessment method (MAD: *Most Apparent Distortion*) which attempts to explicitly model these two separate strategies. Local luminance and contrast masking are used to estimate detection-based perceived distortion in high-quality images, whereas changes in the local statistics of spatial-frequency components are used to estimate appearance-based perceived distortion in low-quality images. We show that a combination of these two measures can perform well in predicting subjective ratings of image quality.

## 1. INTRODUCTION

The ability to quantify the visual quality of an image in a manner that agrees with human vision is a crucial step for any system that processes consumer images. Over the past several decades, research on this front has given rise to a variety of computational methods of image quality assessment. So-called *full-reference* quality assessment methods take as input an original image and a distorted version of that image, and yield as output a prediction of the visual quality of the distorted image relative to the original image. The effectiveness of an image quality assessment method is gauged by examining how well the method can predict ground-truth, human-supplied quality ratings obtained via subjective testing.

The earliest methods of full-reference quality assessment were based primarily on the energy of the distortions. Classical examples include mean-squared error (MSE) and peak-signal-to-noise ratio (PSNR), which operate based on pointwise differences of digital pixels values; and root-mean-squared (RMS) contrast, which operates based on pointwise differences of physical luminances.[1] More recently, methods have been developed based on properties of the human visual system (HVS).[2–20] The vast majority of HVS-based methods employ a "perceptual decomposition" which mimics the local spatial-frequency analysis performed in early vision. This decomposition is typically followed by processing stages which take into account near-threshold psychophysical properties such as contrast sensitivity and visual masking. The most recent HVS-based methods have also begun to model higher-levels of vision.[11,18–20] Another class of methods have recently been proposed which do not explicitly model the stages of vision, but instead operate based on overarching principles of what the HVS is trying to accomplish when viewing a distorted image.[21–24] Overarching principles typically include some form of structural or information extraction, which assumes that a high-quality image is one whose structural content (object boundaries and/or regions of high entropy) most closely matches that of the original image.

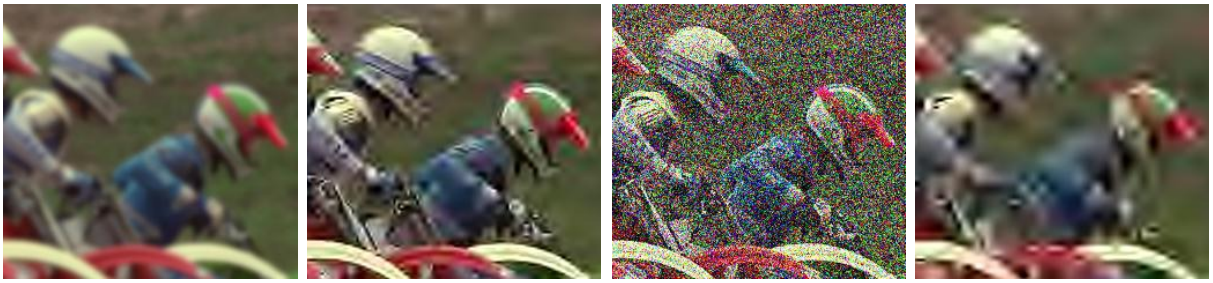
Methods which operate based only on the energy of the distortions, such as MSE and PSNR, are attractive due to their mathematical simplicity. However, these methods have also been shown to be relatively poor predictors of visual quality,[25] particularly when comparing images containing different types of distortions. Methods which take into account properties of the human visual system have demonstrated great success at predicting quality for images containing near-threshold (just-visible) distortions. However, these methods generally perform less well for highly distorted images containing suprathreshold (clearly visible) distortions[26–28] unless properties of

E.C.L.: E-mail: ericcl@okstate.edu; D.M.C.: E-mail: damon.chandler@okstate.edu

Image Quality and System Performance VI, edited by Susan P. Farnand, Frans Gaykema,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7242, 72420S · © 2009 SPIE-IS&T
CCC code: 0277-786X/09/$18 · doi: 10.1117/12.810071

(a) Original image (b) JPEG-2000 compression DMOS = 17.7 (c) White noise DMOS = 23.5

**Figure 1.** When judging the quality of a distorted image containing near-threshold distortions, one tends to rely primarily on visual detection (often using point-by-point comparisons with the original image) in an attempt to locate any visible differences. (a) Close-up of original image *bikes*; (b) close-up of image *bikes* distorted via JPEG-2000 compression; (c) close-up of image *bikes* distorted by additive Gaussian white noise. DMOS values indicate differential mean opinion scores from the LIVE image database.[29]



(a) Gaussian blurring DMOS = 59.0 (b) JPEG-2000 compression DMOS = 67.1 (c) White noise DMOS = 74.6 (d) JPEG-2000 compression DMOS = 82.7

**Figure 2.** When judging the quality of a distorted image containing clearly visible (suprathreshold) distortions, one tends to rely much less on visual detection and much more on overall image appearance in an attempt to recognize image content in the presence of the dominating distortions. (a) Close-up of image *bikes* distorted via Gaussian blurring; (b) and (d) close-up of image *bikes* distorted via JPEG-2000 compression; (c) close-up of image *bikes* distorted by additive Gaussian white noise. DMOS values indicate differential mean opinion scores from the LIVE image database.[29]

suprathreshold vision are also taken into account (e.g., Refs. 19, 20). In contrast, methods based on overarching principles have demonstrated success for images containing suprathreshold distortions. However, because these methods largely ignore properties of early vision, they are not generally suited to high-quality images containing near-threshold distortions.

Despite the clear differences in the way these approaches operate, the vast majority of existing methods share a common thread. Namely, they are rooted in the assumption that when a human determines image quality, the HVS operates via a *single* strategy. For MSE/PSNR, the assumption is that the strategy employed by the HVS is to gauge the intensity of the distortions. For methods based on near-threshold psychophysics, the assumption is that the strategy employed by the HVS is to process the images via a local spatial-frequency decomposition with adjustments for masking, and then to collapse this perceptual decomposition into a final value of quality. For methods based on overarching principles, the common assumption is that the strategy employed by the HVS is to extract local image structure or use natural image statistics to extract the maximal amount of information, and thus quality is determined based on the extent to which this information can be extracted.

In this paper, we advocate an approach to image quality assessment which builds on the strengths of previous approaches, but which operates based on a fundamentally different premise: We assume that the HVS is

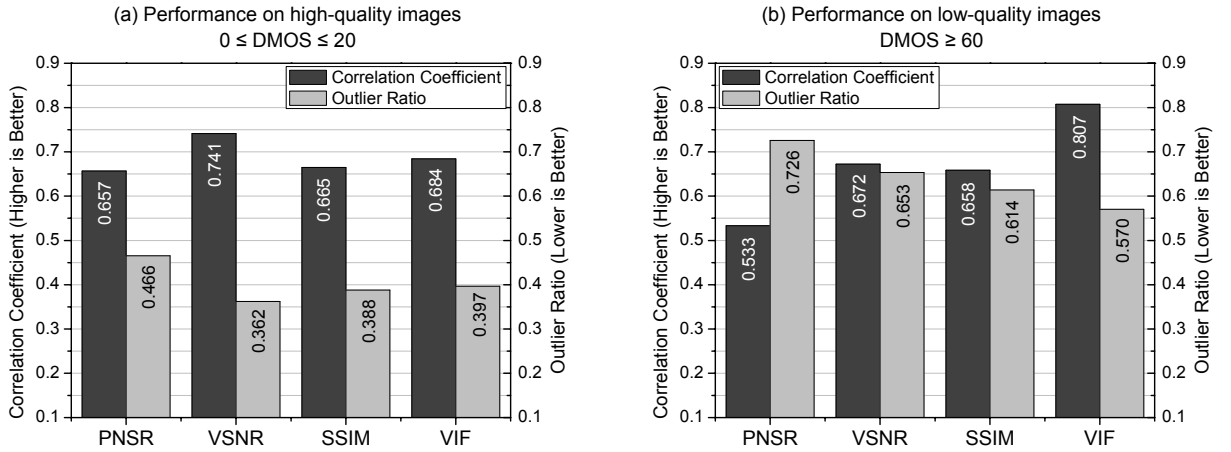



**Figure 3.** Correlation coefficients and outlier ratios for four image quality assessment methods—PNSR, VSNR,[20] SSIM,[21] and VIF[32]—on images from the LIVE image database[29] (see Section 3 for details of the analysis). Notice that VSNR, which takes into account visual masking, performs best for high-quality images; whereas VIF, which operates based on the overarching principle of information extraction, performs best for low-quality images. The fact that the most effective method of assessment varies for high-quality vs. low-quality images suggests that perhaps the HVS employs different strategies of assessment for different levels of distortion.

performing *multiple* strategies when determining quality. Numerous studies have shown the HVS to be a highly adaptive system, with adaptation occurring at multiple levels ranging from single neurons[30] to entire cognitive processes.[31] Thus, even if a human observer is given a fixed, single task of judging image quality, it is reasonable to assume that different strategies might be employed for different conditions (e.g., for different images, for different image regions, and/or for different types and amounts of distortion). Here, we present an image quality assessment method which attempts to explicitly model two strategies employed by the HVS: (1) A detection-based strategy for high-quality images containing near-threshold distortions; and (2) an appearance-based strategy for low-quality images containing clearly suprathreshold distortions.

The need to explicitly model these two separate strategies was motivated by our own experiences when simultaneously judging the qualities of several distorted versions of the same original image. We observed that when viewing and judging the quality of each distorted image, the HVS tends to concentrate on different aspects of the images. Specifically, as shown in Figure 1, some of the distorted images contained just-visible (near-threshold) distortions; these images were consequently judged to be of relatively high quality compared to the original image [Figure 1(a)]. For these higher quality images, because the distortions are not readily visible, our visual system seems to employ a *detection* strategy in an attempt to locate any visible differences. Other distorted images, shown in Figure 2, contained clearly visible (suprathreshold) distortions and were consequently judged to be of lower quality. For these lower quality images, the distortions dominate the overall appearance of each image, and thus visual detection is less applicable. Instead, for these latter images, quality is determined based primarily on our ability to *recognize* image content. Thus, in the high-quality regime, the HVS attempts to look for distortions in the presence of the image; whereas in the low-quality regime, the HVS attempts to look for image content in the presence of the distortions. We argue that these two fundamentally different strategies require two separate computational models.

Another motivation for explicitly modeling two separate strategies can be seen by examining the performances of existing quality assessment methods on high-quality vs. low-quality images. Figure 3 shows the performances of four existing methods of quality assessment on images from the LIVE image database.[29] The results in Figure 3(a) correspond to images which were rated by human observers to be of relatively high visual quality (differential mean-opinion scores between 0 and 20). The results in Figure 3(b) correspond to images which were rated to be of relatively low visual quality (differential mean opinion scores of 60 and above). The four methods reported here are:

1. *PSNR* (Peak Signal-to-Noise Ratio), which operates based only on the energy of the distortions.
2. *VSNR* (Visual Signal-to-Noise-Ratio),[20] an HVS-based method which operates based on low-level masking and mid-level global precedence (the HVS's preference for coarse spatial scales).
3. *SSIM* (Structural SIMilarity),[21] which operates based on the overarching principle that the HVS has evolved to extract structural information from natural images, and therefore a high-quality image is one whose structure most closely matches that of the original.
4. *VIF* (Visual Information Fidelity),[32] which operates under the overarching principle that the HVS has evolved based on the statistical properties of the natural environment, and therefore the visual fidelity of a distorted image can be quantified based on the amount of information it provides about the original image.

Notice from Figure 3(a) that for high-quality images, PSNR, which operates based on the energy of pointwise errors, is as effective as SSIM and VIF which operate based on statistically-based structural/information extraction. VSNR, which also operates based on pointwise errors, outperforms the other three by taking into account contrast sensitivity and luminance and contrast masking. The fact that masking and other properties of early vision are most effective for high-quality images suggests that perhaps the HVS is indeed employing a detection strategy for these types of images. On the other hand, the opposite trend is observed for low-quality images [Figure 3(b)]. For these images, which contain clearly visible distortions, VIF is the most effective, suggesting that near-threshold masking is less important when the distortions are clearly suprathreshold; rather, the HVS is perhaps employing an alternative strategy.* Thus, the fact that the most effective method of assessment varies for high-quality vs. low-quality images suggests that perhaps the HVS employs different strategies of assessment for different levels of distortion.

We have titled our quality assessment method *Most Apparent Distortion* (MAD) to stress the fact that what is most apparent to the human observer—and thus the strategy employed by the HVS—can change depending on the amount of distortion. MAD operates by using both a detection-based model and an appearance-based model. For detection, we employ a simple spatial-domain model of local visual masking which takes into account the contrast sensitivity function and luminance and contrast masking with distortion-type-specific adjustments. Detection-based quality is then estimated based on the mean-squared error (in the lightness domain) between the original and distorted images computed locally for those regions in which the masking models deem the distortions to be visible. For appearance, we employ a model which follows from the texture-analysis literature. The original and distorted images are first decomposed using a log-Gabor filterbank, and the resulting coefficients corresponding to each spatial scale are weighted, with greater weight assigned to coarser scales. Appearance-based quality is then estimated based on the absolute difference between low-level statistics (mean, variance, skewness, and kurtosis) measured for the weighted coefficients of the the original and distorted images. The overall quality of the distorted image is computed by taking a weighted geometric mean of the detection-based and appearance-based qualities, where the weight is determined based on the amount of distortion. For highly distorted images, greater weight is given to the appearance-based quality, whereas for images containing near-threshold distortions, greater weight is given to the detection-based quality.

This paper is organized as follows. Section 2 describes the details of the MAD algorithm. Section 3 provides an analysis of MAD's performance in predicting subjective quality ratings from both the LIVE image database and from a new image database (CSIQ) which we have recently made available to the research community. General conclusions are provided in Section 4.

## 2. ALGORITHM

This section describes the details of the MAD algorithm. First, we describe a method for quantifying perceived distortion in images containing near-threshold distortions (relatively high-quality images); this first method is used to model visual detection. Next, we describe a method for quantifying perceived distortion in images

*VSNR still performs well in the low-quality range because it employs a wavelet-based model of global precedence (scale-space weighting) in addition to modeling low-level masking; this global precedence adjustment is negligible for high-quality images.

containing suprathreshold distortion (relatively low-quality images); this latter method is used to model image appearance. Finally, we describe a technique used to combine the two perceived distortions into a single estimate of overall perceived distortion.

## 2.1. A Detection-Based Strategy for High-Quality Images

When viewing a high-quality image, we argue that the HVS attempts to look for distortions in the presence of the image. To approximate this visual detection strategy, we combine a spatial-domain model of local masking with local mean-squared error measured in the perceived luminance (lightness) domain. First, low-level psychophysical properties such as the spatial contrast sensitivity function, the nonlinear perception of luminance, and luminance and contrast masking are used to compute a map denoting the locations of visible distortions. Next, this visibility map is used to compute a visibility-weighted local MSE map. Finally, the perceived distortion is estimated by collapsing the visibility-weighted local MSE map (via the $L_2$ norm) into a single scalar value.

### 2.1.1. Step 1: Compute Locations at which the Distortions are Visible

#### Perceived Luminance

To account for the nonlinear relationship between digital pixel values and physical luminances of typical display media, the pixels of the original and distorted images are first converted to luminances (in cd/m$^2$) via

$$L = (b + k \times I)^{\gamma} \tag{1}$$

where $I$ is the digital pixel value, and where the parameters $b$, $k$, and $\gamma$ are constants specific to the device on which the images are to be displayed. For 8-bit pixel values and an sRGB display,[33] the values for these parameters are given by $b = 0$, $k = 0.02874$, and $\gamma = 2.2$.

Next, we take into account the nonlinear HVS response to luminance by converting luminances into *perceived* luminances (relative lightness) via

$$L^* = \sqrt[3]{L} \tag{2}$$

where $L^*$ denotes the relative lightness and is proportional to the lightness used in the CIELAB color space. Equation (2) is used for both the original and reference images to yield the perceived-luminance images $L^*_{org}$ and $L^*_{dst}$, respectively. We then define the error image as $L^*_{err} = L^*_{org} - L^*_{dst}$.

#### Contrast Sensitivity Function

We account for variations in (near-threshold) sensitivity to spatial frequency by using a model of the contrast sensitivity function (CSF) from Ref. 18. This CSF has a peak sensitivity of 6 cycles per degree of visual angle, and a rolloff of 3 dB along the diagonals. The CSF is applied by filtering both the original image and the error image, where the filtering is performed in the frequency domain via

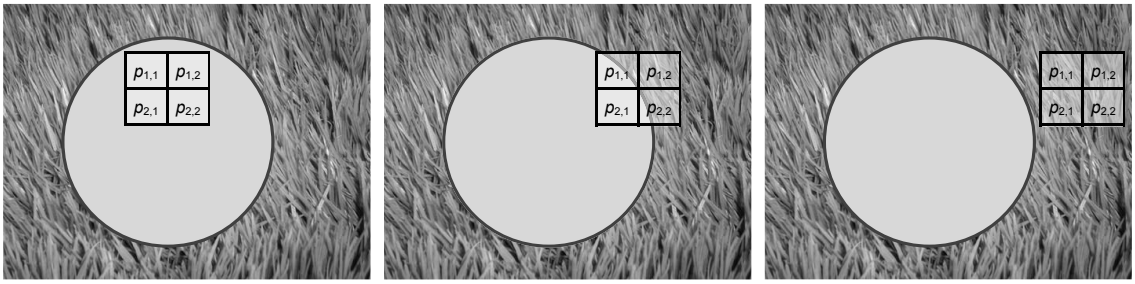
$$I'_x = F^{-1}\left[CSF \times F\left[L^*_x\right]\right] \tag{3}$$

where $F[\cdot]$ and $F^{-1}[\cdot]$ denote the discrete Fourier transform (DFT) and inverse DFT, respectively, and where $L^*_x$ denotes the original or error image. At this point, $I'_{org}$ and $I'_{err}$ represent the original and distorted images with values that are linearly proportional to both perceived luminance and perceived contrast. $I'_{err}$ can be considered to be the distortions in the image that the HVS could detect if the distortions were viewed against a uniform background (i.e., no masking) rather than being viewed against the image.

#### Contrast Masking

To account for the fact that the presence of an image can reduce the detectability of distortions, we employ a simple spatial-domain measure of contrast masking. First, a local contrast map is computed for the original image by dividing $I'_{org}$ into $16 \times 16$ blocks (with 75% overlap between neighboring blocks), and then measuring the RMS contrast $C_{org}(p)$ for each block $p$ via

$$C_{org}(p) = \tilde{\sigma}_{org}(p)/\mu_{org}(p) \tag{4}$$




(a) Contrast in block is measured to be low

(b) Contrast in block is measured to be low due to low contrast in $p_{2,1}$

(c) Contrast in block is measured to be high

**Figure 4.** Example of measuring the local (block-based) contrast in an image containing a low-contrast object (here, a solid disc) placed upon a high-contrast background, as is often found in natural images. In (a), the contrast of the block is low. In (c), the contrast of the block is high. In (b), the contrast of the block would normally be higher than the contrast measured in (a), despite the fact that this object border is not effective at masking.[34] Here, we overcome this limitation by taking the effective local contrast to be the minimum contrast of all four sub-blocks. Thus, in (b), since sub-block $p_{2,1}$ has the lowest contrast, the effective local contrast is measured to be just as low as that measured in (a).

where $\mu_{org}(p)$ denotes the mean of block $p$ of $I'_{org}$, and where $\tilde{\sigma}_{org}(p)$ denotes the modified standard deviation of block $p$ of $I'_{org}$ given by

$$\tilde{\sigma}_{org}(p) = \min\left[\sigma(p_{1,1}), \sigma(p_{1,2}), \sigma(p_{2,1}), \sigma(p_{2,2})\right] \tag{5}$$

where the four standard deviations of the sub-blocks are denoted by $\sigma(p_{x,y})$ ($x, y$ correspond to the position of the sub-block as shown in Figure 4). This modified standard deviation is used to compensate for the fact that edges and object boundaries are not effective at masking, despite the fact that these regions exhibit high contrast.[34] As shown in Figure 4, by defining the effective contrast based on the minimum contrast of each block's four sub-blocks, the resulting contrast around edges/object borders is measured to be relatively low.

Whereas $C_{org}(p)$ is a measure of the local RMS contrast in the original image (and thus a measure of the relative masking capability of each region), $C_{org}(p)$ is independent of the distortions. Accordingly, we next compute a local contrast map for the error image to account for the spatial distribution of the distortions in the distorted image. $I'_{err}$ is divided into $16 \times 16$ blocks (with 75% overlap between blocks), and then the RMS contrast $C_{err}(p)$ for each block $p$ is computed via

$$C_{err}(p) = \begin{cases} \sigma_{err}(p)/\mu_{org}(p), & \text{if } \mu_{org}(p) > 0.9 \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

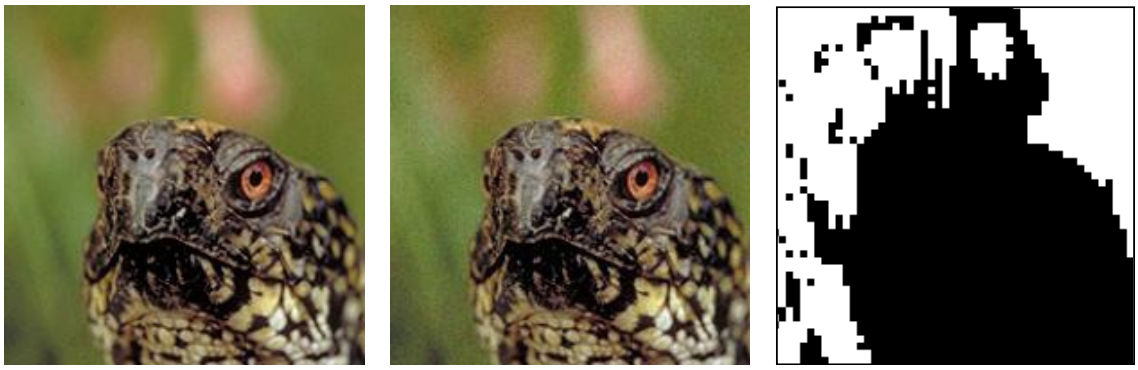
where $\sigma_{err}(p)$ denotes the standard deviation of block $p$ of $I'_{err}$. Here, we employ a lightness threshold of 0.9 to account for the fact the HVS is insensitive to changes in extremely dark regions.

Finally, we use the two local contrast maps $\{C_{org}(p)\}$ and $\{C_{err}(p)\}$ to compute a local distortion visibility map via

$$Visib(p) = \begin{cases} 1, & \text{if } C_{err}(p) > \frac{3}{4} \times C_{org}(p) \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Thus, $Visib(p)$ is a binary value which denotes whether the distortions in block $p$ are visible in the distorted image. We consider the distortions visible if $C_{err}(p) > \frac{3}{4} \times C_{org}(p)$, where the threshold value of $\frac{3}{4}$ was chosen experimentally by visually inspecting a number of generated masking maps and distorted images.

Figure 5 depicts an original image *turtle*, a distorted version of that image, and the resulting visibility map $\{Visib(p)\}$. Notice that the masking imposed by the turtle's skin is well captured by the map. Although this spatial-domain model of masking is certainly not a proper model of how masking is effected in the HVS, it provides a reasonable tradeoff between accuracy and computational efficiency. In particular, standard deviation can be approximated efficiently using a fast, separable convolution.



(a) Original image (b) Distorted image (c) Map of visible distortions

**Figure 5.** (a) Original image *turtle* from the CSIQ image database.[35] (b) Image *turtle* distorted via additive Gaussian white noise. (c) Binary map in which white blocks denote the locations at which the distortions in (b) are visible.

### 2.1.2. Step 2: Combine the Visibility Map with Local Errors

After the map of visible locations is created, we use visibility-weighted local MSE measured in the lightness domain to determine perceived distortion. Although MSE on its own is not always a veridical indicator of perceived distortion, as we will demonstrate in Section 3, when MSE is computed locally and combined with a visibility map, it can perform well in the high-quality (low-distortion) regime.

We define the perceived distortion in the high-quality regime as:

$$PD_{high} = \frac{1}{N}\sqrt{\sum_{p} Visib(p) \times LMSE(p)^2} \tag{8}$$

where the summation is over all blocks and $N$ is the total number of blocks. The quantity $LMSE(p)$ is the local MSE computed for each $16 \times 16$ block $p$ via

$$LMSE(p) = \frac{1}{16^2}\sum_{i,j \in N_p} I'_{err}(i,j)^2 \tag{9}$$

where $N_p$ is the set of pixels inside block $p$.

Equation (8) collapses the visible LMSE into a single value using the $L_2$ norm, which is a reasonable approximation of visual summation of distortion in natural images.[36] A value of $PD_{high} = 0$, indicates that the distortions in the distorted image are not visible. Increasing values of $PD_{high}$ denote increasing perceived distortion and thus decreasing visual quality.

Figure 6 shows an example of the maps involved in computing $PD_{high}$ for an image from LIVE image database distorted via JPEG compression. The original and distorted images are shown in Figures 6(a) and (b), respectively. Notice that the only visible distortions occur in the sky, particularly along the boundary between the buildings and skyline. Additionally, the highway lane markers show distinct blocking artifacts.

Figures 6(c), (d), and (e) show the computed visibility map, local MSE map, and visibility-weighted local MSE map $[Visib(p) \times LMSE(p)]$, respectively. Notice from Figure 6(c) that the visibility map does well at capturing the visible artifacts (though there is some noise, overall, the sky and highway distortions are well detected). Notice from Figure 6(d) that because local MSE does not take into account masking, the local MSE map indicates that the greatest distortions occur in busy areas, whereas the distortions in these regions are not visible. Finally, notice from Figure 6(e) that the visibility-weighted local MSE map performs well at indicating both the locations and perceived intensities of the visible distortions. In particular, this latter map captures the fact that that some artifacts are visible in the sky but do not really degrade the visual quality, and that the most annoying artifacts occur in the street and along the skyline.

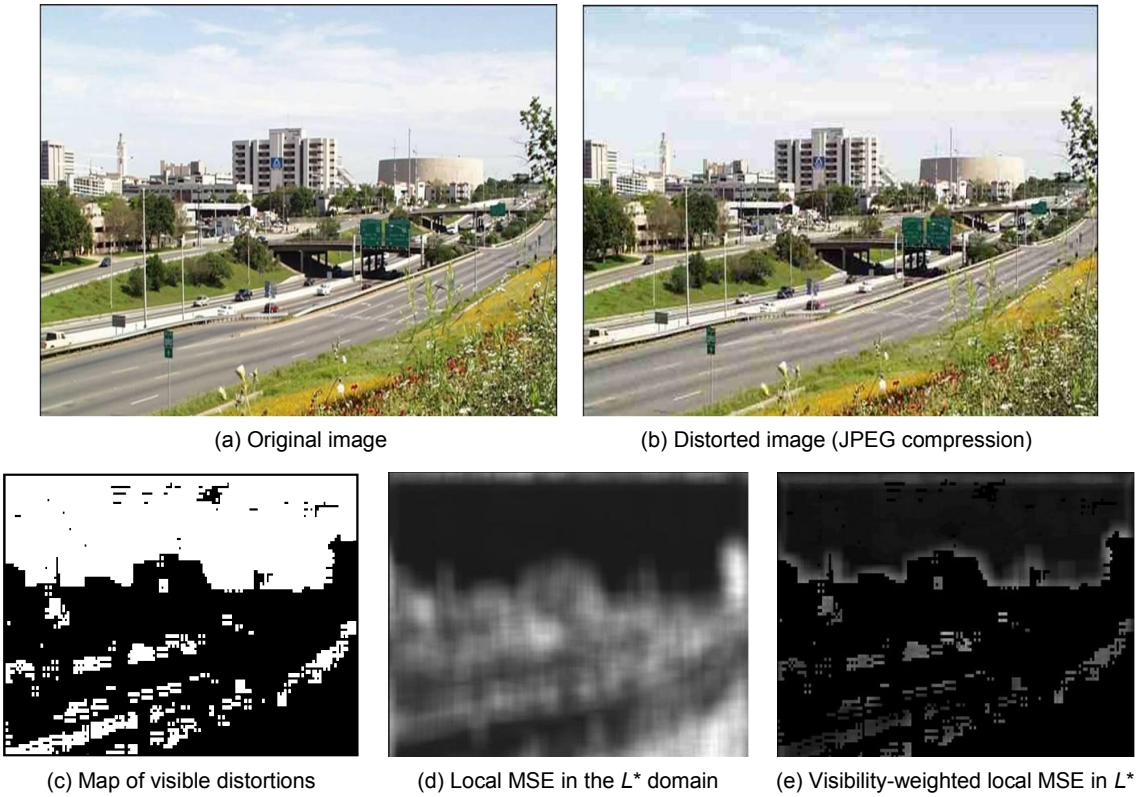



**Figure 6.** Perceived distortion in the high-quality regime is largely determined based on masking. Given an original image (a), and a distorted image (b), a map denoting the locations of visible distortions is computed (c), which is then multiplied by local MSE measured in the lightness domain (d). The resulting visibility-weighted local MSE (d) is used to determine $PD_{high}$.

## 2.2. An Appearance-Based Strategy for Low-Quality Images

In the low-quality regime, we argue that visual masking is of lesser importance to our perception of quality; rather, when the distortions are highly suprathreshold, perceived distortion is better modeled by quantifying the extent to which the distortions degrade the appearance of the image's subject matter. Thus, the HVS switches from a strategy of detecting distortions in the high-quality regime to a strategy of judging image appearance in the low-quality regime.

Here, we propose using a biologically motivated model of appearance based on local statistics of multi-scale log-Gabor filter responses. This method has long been used by the computer vision community to classify the appearance of textures,[37] and we showed in Ref. 38 that these statistics are good indicators of how camouflaged an animal *appears* in its natural environment. Furthermore, previous researchers have shown that simple cells in primary visual cortex are well-modeled using log-Gabors,[39,40] and that for textures, changes in the statistics of log-Gabor filter responses are more visually apparent than changes in pixel statistics.[41]

### 2.2.1. Step 1: Apply a Log-Gabor Decomposition

The original and distorted images are first decomposed into a set of subbands using a log-Gabor filterbank. The filtering used to obtain the subbands is performed in the frequency domain by computing the inverse DFT of the product of the image's DFT with the following 2D frequency response:

$$G_{s,o}(f,\theta) = \exp\left(-\frac{(\log f/f_s)^2}{2\left(\log \sigma_s/f_s\right)^2}\right) \times \exp\left(-\frac{(\theta-\mu_o)^2}{2\sigma_o^2}\right) \tag{10}$$

where $G_{s,o}$ is the filter defined at spatial scale $s$ and orientation $o$. The parameter $f_s$ denotes the center frequency of the scale, and $\sigma_s/f_s$ is a measure of the bandwidth. The parameters $\mu_o$ is the center orientation of the filter and $\sigma_o$ is the angular spread. Because the log-Gabor filters have been shown to approximate cortical responses in primary visual cortex,[40] the parameters $f_s$, $\sigma_s$, $\mu_o$, and $\sigma_o$ can be selected to match corresponding estimates obtained from the mammalian visual system. Here, we have selected values reported in Ref. 42.

The log-Gabor decomposition is computed at five scales (3, 6, 16, 32, and 64 pixels) and four orientations (0°, 45°, 90°, and 135°). The value of $\sigma_s/f_s$ is set to 0.65, which translates to each filter spanning approximately 1.5 octaves. The bandwidth of each subband is chosen to minimize overlap between bands, but still uniformly cover most of the frequency spectrum (as in the mammalian visual system).

Because the frequency responses of the filters defined by Equation (10) are non-zero only for positive frequencies (and thus not conjugate symmetric), multiplication of each image's DFT by the frequency response is equivalent to convolving each image in the space domain with both even- and odd-symmetric filters; i.e., the inverse DFT yields complex subband coefficients. We collapse the even and odd filter outputs (real and imaginary components) into a single magnitude, thus yielding 20 subbands (5 scales × 4 orientations) containing only positive coefficients.

#### 2.2.2. Step 2: Compare Subband Statistics

After computing the 20 subbands each for the original and distorted images, the local subband statistics of the original image are compared with the corresponding local subband statistics of the distorted image to define a local statistical difference map $\{\eta(p)\}$. Specifically, for each each $16\times16$ block (with 75% overlap between blocks), the difference in standard deviation, skewness, and kurtosis of the block's corresponding subband coefficients is computed via

$$\eta(p) = \sum_{s,o} w_s \left( |\sigma_{s,o}^{org} - \sigma_{s,o}^{dst}| + 2|\xi_{s,o}^{org} - \xi_{s,o}^{dst}| + |\kappa_{s,o}^{org} - \kappa_{s,o}^{dst}| \right) \tag{11}$$

where $\sigma_{s,o}$, $\xi_{s,o}$, and $\kappa_{s,o}$ denote, respectively, the standard deviation, skewness, and kurtosis of the $16 \times 16$ subband coefficients corresponding to scale $s$ and orientation $o$ (and corresponding in location to block $p$). Note that the skewness difference is multiplied by a factor of 2 to bring it to approximately the same scale as the $\sigma$ and $\kappa$ differences. Also note that scale-specific weights, $w_s$, are used to account for the HVS's preference for coarse scales over fine scales (cf Ref. 43). The values chosen for $w_s$ for the finest to coarsest scales are 0.5, 0.75, 1, 5, and 6, which were selected to yield the best performance on the A57 database.[44] Using these weights, the finest scale contributes approximately 3.8% to $\eta(p)$, and coarser scales contribute 5.6%, 7.5%, 37.7%, and 45.2%, respectively.[†]

The statistics of the log-Gabor filter outputs have been widely used to define visual appearance and texture. Specifically, changes in standard deviation, skewness, and kurtosis have been shown to be good indications of discriminable texture appearance.[41] A change in the standard deviation of log-Gabor subband coefficients means that the outputs of certain log-Gabor filters change in intensity. When these changes are computed on a block-by-block basis, we are also able to approximately locate where in the image the filter outputs change. For instance, blurring dramatically changes the histogram of the log-Gabor outputs: (1) At high frequency, the histogram of filter outputs become more peaked with smaller standard deviation. (2) Additionally, if the outputs were skewed to one side before blurring, the histogram is likely to be more symmetric afterward. Compression artifacts would have different but measurable changes in the log-Gabor statistics. These changes in local subband statistics can approximate the perceived distortion of local image structure. It is important to note that we are not saying the HVS performs statistical comparisons, but that these statistics do well at approximating what the HVS defines as appearance.

The final scalar value of perceived distortion in the low-quality regime is given by

$$PD_{low} = \frac{1}{N} \sqrt{\sum_{p} \eta(p)^2} \tag{12}$$

[†]The proper values of $w_s$ remain an area of future research. Adjustment of $w_s$ only marginally affects the performance of the low-quality prediction.

(a) Original image (b) Distorted image (JPEG compression) (c) Statistical difference map

**Figure 7.** Perceived distortion in the low-quality regime is largely determined based on changes in the appearance of the image's subject matter. Given an original image (a), and a distorted image (b), a map of local changes in log-Gabor filter-response statistics is computed (c). This statistical difference map is used to determine $PD_{low}$.

where the summation is over all blocks and $N$ is the total number of blocks. A value of $PD_{low} = 0$ denotes no perceived distortion and increasing values of $PD_{low}$ denote increasing perceived distortion and thus decreasing visual quality.

Figure 7 shows the resulting statistical difference map $\{\eta(p)\}$ used in computing $PD_{low}$ for an image *caps* from the LIVE image database distorted via JPEG compression. The original and distorted images are shown in Figures 7(a) and (b), respectively. Notice that the most disturbing artifacts manifest as differences in the appearances of the sky, the edges of the wall, and the shadows of the caps. As shown in Figure 7(c), the statistical difference map succeeds at capturing these changes in visual appearance.

### 2.3. Combining the Two Strategies

The previous sections presented two measures of perceived distortion: $PD_{high}$ designed for high-quality images (i.e., images containing near-threshold distortions), and $PD_{low}$ designed for low-quality images (i.e., images containing clearly suprathreshold distortions). In this section, we describe how these two measures are combined to yield an overall measure of perceived distortion, applicable for all images, based upon how apparently distorted the image appears.

High quality images should obtain their rating mostly from $PD_{high}$ and low quality images from $PD_{low}$. We hypothesize that in the transition between high and low quality assessment, observers use a mixture of strategies. This hypothesis makes sense for images in which some regions may appear to be of high quality, while other regions may contain clearly suprathreshold distortions. Images compressed with JPEG and JPEG-2000, for example, fall into this category. Figure 8 shows an example. In comparison to the original image shown in Figure 8(a), the distorted image shown in Figure 8(b) contains a mixture of what one would consider high-quality and low-quality regions. The left sidewall of the bridge and the mountain in the background exhibit statistical appearance changes (severe blurring), while the high-contrast trees and brighter areas of the sky largely mask the distortions. In addition, the middle slats of the bridge look somewhat normal, whereas the more distant slats appear clearly degraded.

To capture the interacting strategies, we propose using a weighted geometric mean of $PD_{high}$ and $PD_{low}$ given by

$$MAD = (PD_{high})^{\alpha} (PD_{low})^{1-\alpha} \tag{13}$$

where $MAD \in [0, \infty]$ denotes the overall perceived distortion. The weight $\alpha \in [0, 1]$ is chosen based on the (predicted) overall level of distortion. For low levels of distortion, MAD should obtain its value mostly from $PD_{high}$ (i.e., $\alpha$ should approach a value of 1). For high levels of distortion, MAD should obtain its value mostly from $PD_{low}$ ($\alpha$ should approach a value of 0).

(a) Original image

(b) Distorted image (JPEG-2000) containing both high-quality and low-quality regions

**Figure 8.** Image from the CSIQ image database[35] containing regions of both high quality (trees and nearer slats of the bridge) and low quality (the sidewall and distant slats of the bridge, the mountain in the background). (a) Original image *bridge*. (b) Distorted version of *bridge* containing JPEG-2000 compression artifacts.

Although the optimal technique of selecting $\alpha$ remains an area of future research, we have found that selecting $\alpha$ based on $PD_{high}$ can yield good performance. Here, $\alpha$ is computed via

$$\alpha = \frac{1}{1 + \beta_1 \left(PD_{high}\right)^{\beta_2}} \tag{14}$$

where $\beta_1$ and $\beta_2$ are free parameters. For the A57 database,[44] the optimal values for these parameters are $\beta_1 = 0.135$ and $\beta_2 = 0.869$.

## 3. RESULTS

In this section, the performance of MAD is analyzed in terms of its ability to predict subjective ratings of image quality. To assess its predictive performance, MAD was applied to two databases of subjective image quality: (1) The LIVE database,[45] and (2) a new database of subjective quality entitled the *Categorical Subjective Image Quality* (CSIQ) database.[35, 46]

The LIVE database contains 29 original images, 26 to 29 distorted versions of each original image, and subjective ratings of quality for each distorted image (differential mean opinion scores, DMOS values). The distortions used in LIVE are: Gaussian blurring, additive white noise, JPEG compression, JPEG-2000 compression, and simulated data packet loss of transmitted JPEG-2000-compressed images. The DMOS values were provided as part of the LIVE image database; they were not experimentally determined nor verified in the current study.

The current preliminary version of the CSIQ database consists of ten original images distorted using six different types of distortions at five different levels of distortion. The distortions used in CSIQ are: JPEG compression, JPEG-2000 compression, global contrast decrements, additive pink Gaussian noise, additive white Gaussian noise, and Gaussian blurring. This results in 10 original images and 300 distorted versions of the images. Details of the CSIQ database can be seen in the online supplement.[46] We are working to expand the database, however, the preliminary findings are used to show the extent to which MAD generalizes to novel distortions.

### 3.1. Performance Measures

Before evaluating the performance of a particular quality assessment method on a particular database, it is customary to apply a logistic transform to the predicted ratings so as to bring the predictions on the same scale as the DMOS values, and to attempt to obtain a linear relationship between the predictions and DMOS values. The logistic transform recommended by the Video Quality Experts Group[47] is a four parameter sigmoid given by

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp \frac{x - \tau_3}{\tau_4}} + \tau_2. \tag{15}$$

The parameters $\tau_1, \tau_2, \tau_3$, and $\tau_4$ are chosen to minimize the MSE between the predicted values and the DMOS values. The logistic transform is monotonic and was chosen mainly for its ability to facilitate the use of various performance measures.

The most common measures of predictive performance are the Pearson correlation coefficient (CC) and Spearman rank-order correlation coefficient (SROCC). CC and SROCC are measures of how well an algorithm's predictions correlate with the raw DMOS values and how well an algorithm predicts the relative ordering of the distorted images, respectively.

There are also two measures of performance more specific to image quality prediction: The outlier ratio[48] $R_{out}$, and the outlier distance $d_{out}$. These two measures attempt to account for the inherent variation in human subjective ratings of quality. For instance, if the perceived rating of a particular image has large variation between observers, then the *average* DMOS rating is not necessarily a good indication of what the algorithm should predict. Instead, some leeway should be given around DMOS values associated with the variability of observers. With this in mind, the outlier ratio is defined as[48]

$$R_{out} = \frac{N_{false}}{N_{total}} \tag{16}$$

where $N_{false}$ is the number of predictions outside two standard deviations, $2\sigma_{dmos}$, of the average DMOS and $N_{total}$ is the total number of predicted DMOS values. $R_{out}$ indicates how often an algorithm predicts subjective quality within a given range. The range of $2\sigma_{dmos}$ was chosen because it contains 95% of all the subjective quality scores for a given image.

In addition to knowing if a predicted rating is an outlier, it is also informative to know how far outside of the error bars ($\pm 2\sigma_{dmos}$) the outlier lies. To quantify this, we propose a new measure, termed the *outlier distance*, which is the distance from an outlier to the closest error bar. The outlier distance $d_{out}$ is defined as

$$d_{out} = \sum_{x \in X_{false}} \min \left| f(x) - [DMOS(x) + 2\sigma_{dmos}(x)] \right|, \left| f(x) - [DMOS(x) - 2\sigma_{dmos}(x)] \right| \tag{17}$$

where $f(x)$ is the predicted DMOS rating as defined in Equation (15) and $X_{false}$ is the set of all predicted ratings outside $2\sigma_{dmos}$.

### 3.2. Performance at High and Low Quality

Before examining the performance of MAD on all images from each database, it is informative to examine how the separate strategies of MAD perform on high-quality vs. low-quality images (i.e., using just $PD_{high}$ and $PD_{low}$ on subsets of images containing only high-quality or only low-quality images). Table 1 (*left*) shows the performances of various quality assessment algorithms on high-quality images from the LIVE database ($0 \leq$ DMOS $\leq 20$). Table 1 (*right*) shows the performances on low-quality images from the LIVE database (DMOS $\geq 60$). Also listed in Table 1 are results from PSNR, SSIM,[21] VSNR,[44] and VIF.[32]

Notice that indeed $PD_{high}$ performs the best on high-quality images and performs poorly on low-quality images. Similarly, $PD_{low}$ performs the best on low-quality images and performs poorly on high-quality images. These results further support the use of separate strategies for different quality regimes.

**Table 1.** Performances of $PD_{high}$, $PD_{low}$, and other quality assessment algorithms on images from the LIVE database rated to be of high quality (DMOS $\leq$ 20) and images rated to be of low quality (DMOS $\geq$ 60).

| | **LIVE High-Quality Images (DMOS $\leq$ 20)** | | | | | | **LIVE Low-Quality Images (DMOS $\geq$ 60)** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *PSNR* | *SSIM* | *VSNR* | *VIF* | $PD_{high}$ | $PD_{low}$ | *PSNR* | *SSIM* | *VSNR* | *VIF* | $PD_{high}$ | $PD_{low}$ |
| *CC* | 0.656 | 0.664 | 0.741 | 0.684 | **0.780** | 0.235 | 0.533 | 0.658 | 0.672 | 0.807 | 0.527 | **0.885** |
| $R_{out}$ | 0.465 | 0.387 | 0.362 | 0.396 | **0.301** | 0.612 | 0.725 | 0.613 | 0.653 | 0.570 | 0.722 | **0.400** |
| $d_{out}$ | 308 | 327 | 212 | 303 | **210** | 510 | 1440 | 1134 | 1099 | 726 | 1452 | **429** |

**Table 2.** Performances of MAD and other quality assessment algorithms on all images from the LIVE and CSIQ databases.

| | **LIVE, All Images** | | | | | **CSIQ, All Images** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *PSNR* | *SSIM* | *VSNR* | *VIF* | *MAD* | *PSNR* | *SSIM* | *VSNR* | *VIF* | *MAD* |
| *CC* | 0.8707 | 0.9378 | 0.9233 | 0.9595 | **0.9695** | 0.8455 | 0.8893 | 0.8472 | 0.9079 | **0.9487** |
| *SROCC* | 0.8763 | 0.9473 | 0.9278 | 0.9633 | **0.9703** | 0.8428 | 0.9019 | 0.8577 | 0.9063 | **0.9469** |
| $R_{out}$ | 0.682 | 0.592 | 0.588 | 0.546 | **0.424** | 0.356 | 0.305 | 0.282 | 0.339 | **0.235** |
| $d_{out}$ | 4943.3 | 2814.1 | 3246.8 | 1890.4 | **1300.8** | 782.3 | 497.1 | 737.9 | 577.8 | **330.8** |

### 3.3. Overall Performance

Table 2 shows the performance of MAD and other quality assessment algorithms on all images from the LIVE database and all images from the CSIQ database. For images from the LIVE image database, MAD demonstrates improved predictive performance over PSNR, SSIM, VSNR, and VIF with respect to all four performance measures (CC, ROCC, $R_{out}$, and $d_{out}$).

The results on CSIQ are similar to those on LIVE, with MAD demonstrating the best performance. Notice that MAD generalizes well to the unique distortions in CSIQ (additive pink noise and contrast degradations). It is of note that the correlations on the CSIQ database are markedly lower than on LIVE. We believe this is due to the new distortions in CSIQ that are not present in LIVE. None of the algorithms reported here (including MAD) were constructed to handle these types of distortions. Even still, MAD appears to generalize well to these new distortions.

Scatterplots of DMOS versus each quality assessment algorithm's predictions can be seen in Figure 9 for both LIVE and CSIQ. Observe that MAD demonstrates the tightest fit on both databases. For CSIQ, also notice that the residuals of many of the algorithms appear to be heteroscedastic (non-Gaussian residuals), indicating that CC should not be the primary means of comparison. In that respect, the considerably higher CC achieved by MAD could, in part, be attributed to the homoscedasticity of its residuals. However, MAD still exhibits markedly lower outlier ratios and outlier distances than the other algorithms.

### 3.4. Statistical Significance

To establish statistical significance, we compare each algorithm's residuals (errors in predictions). If the residuals are Gaussian, an $F$-test can be applied to determine the probability that the residuals are drawn from two different distributions and are thus statistically different; smaller residual variance denotes a better prediction. Note that if the residuals are not Gaussian, the test for significance is considerably more difficult and often inconclusive. One formal test of Gaussianity is the Jarque-Bera (JB) statistic.[49] If the JB statistic of a dataset is sufficiently small, one can assume that the data follow a Gaussian distribution. Larger values of the JB statistic denote larger deviations from Gaussianity.

Table 3 shows the summary for overall statistical performance of each algorithm on LIVE and CSIQ. An $F$-test was performed between each pair of algorithms. A "-1" denotes that the algorithm in the row has statistically larger residuals than the algorithm in the column with confidence greater than 99%. A "1" denotes the opposite. A "0" denotes that there is statistically no difference between the residuals of the two predictions. The table

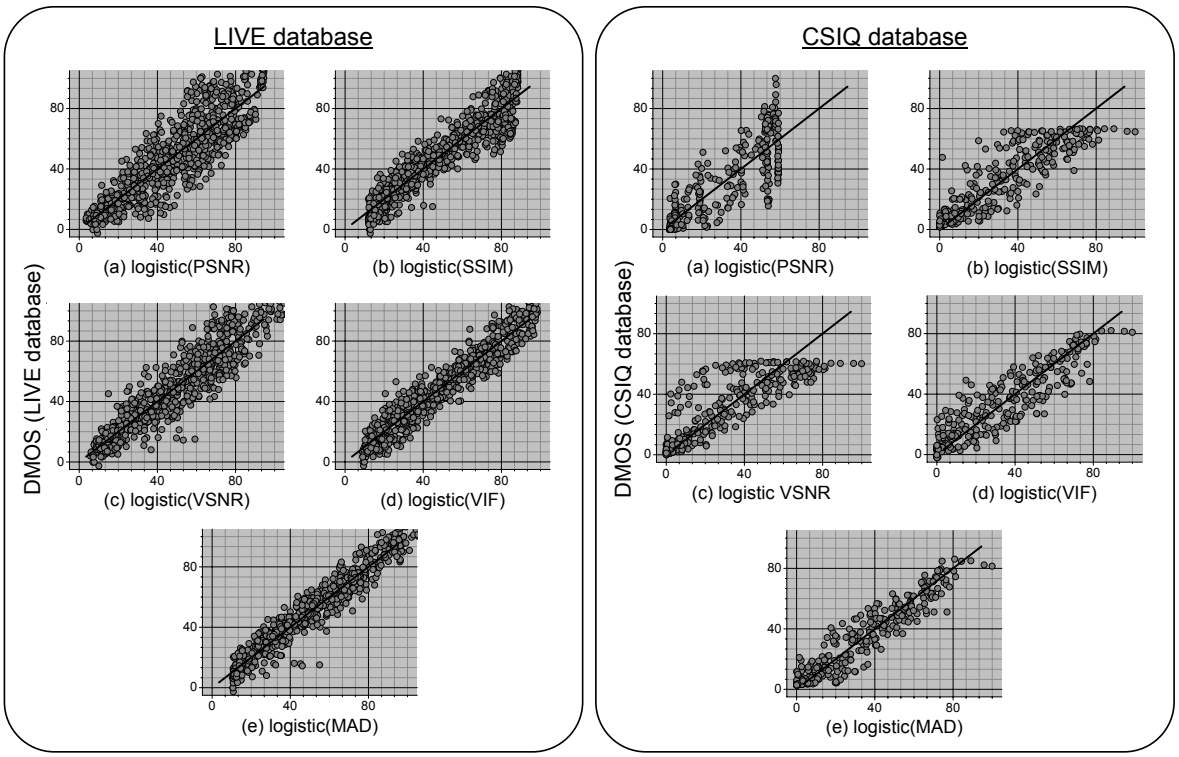



**Figure 9.** The overall logistically transformed fits for PSNR, SSIM, VSNR, VIF, and MAD on the LIVE and CSIQ image databases.

**Table 3.** Statistical significance relationships between the algorithms on the LIVE and CSIQ databases. A "1" denotes that the algorithm in the row has statistically smaller residuals with confidence greater then 99% than the algorithm in the column. A "-1" denotes statistically larger residuals. A "0" denotes that there is no statistical difference. Also shown are the measures of Gaussianity. Italicized JB-test entries denote that the residuals can be deemed gaussian with 95% confidence.

| | **LIVE, Statistical Significance** | | | | | **CSIQ, Statistical Significance** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *PSNR* | *SSIM* | *VSNR* | *VIF* | *MAD* | *PSNR* | *SSIM* | *VSNR* | *VIF* | *MAD* |
| *PSNR* | 0 | -1 | -1 | -1 | -1 | 0 | -1 | 0 | -1 | -1 |
| *SSIM* | 1 | 0 | 1 | -1 | -1 | 1 | 0 | 1 | 0 | -1 |
| *VSNR* | 1 | -1 | 0 | -1 | -1 | 0 | -1 | 0 | -1 | -1 |
| *VIF* | 1 | 1 | 1 | 0 | -1 | 1 | 0 | 1 | 0 | -1 |
| *MAD* | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| *JB Stat* | 11.768 | *2.583* | 20.011 | *4.843* | 250.470 | *0.414* | 17.619 | 24.255 | *0.652* | *2.423* |

also contains the JB statistic measure of Gaussianity. Notice that for both databases, the $F$-tests demonstrate that MAD has statistically smaller residuals than the other algorithms.[‡]

## 4. CONCLUSIONS

This paper presented a new method of image quality assessment which operates under the premise that the HVS performs two distinct strategies when assessing image quality. For high-quality images, because the distortions

[‡]On the LIVE database, MAD is also deemed as the most non-Gaussian of all the algorithms as denoted by the JB statistic. The high JB value is attributable to four outliers (see Figure 9(e) in the LIVE database group). When these four outliers are excluded, MAD's residuals are indeed deemed Gaussian.

are not readily visible, our visual system seems to employ a detection strategy in an attempt to locate any visible differences. For low-quality images, because the distortions tend to dominate the image's overall appearance, visual detection is less applicable, rather, quality is determined based primarily on our ability to recognize image content. Thus, in the high-quality regime, the HVS attempts to look for distortions in the presence of the image; whereas in the low-quality regime, the HVS attempts to look for image content in the presence of the distortions. We argue that these two fundamentally different strategies require two separate computational models.

Accordingly, two separate computational measures of perceived distortion were presented. The first measure, designed for high-quality images, assesses perceived distortion by taking into account contrast sensitivity, and local luminance and contrast masking. The perceived distortion is computed via a visibility-weighted local error measurement computed in the lightness domain ($L^*$). The second measure, designed for low-quality images, assesses perceived distortion based on the changes in local statistics between the subbands of the original image and the subbands of the distorted image. An overall measure of perceived distortion is computed via a weighted geometric mean of the high-quality and low-quality measures, where the weight is determined based on the estimated level of distortion.

The proposed MAD measure was shown to perform well on images from both the LIVE image database[29] and the CSIQ image database.[35] Statistically significant improvements in predicting subjective ratings were achieved in comparison to a variety of existing algorithms. Some notable limitations of MAD include its relatively high computational complexity and memory footprint attributable to the log-Gabor decomposition required for the low-quality measure. In addition, MAD is blind to color-only distortion and has not yet been tested over a range of viewing distances. We are currently in the process of refining the masking model and the log-Gabor decomposition to better take into account viewing distance. We are also exploring extensions of this work for color images and video quality assessment.

## REFERENCES


1. B. Moulden, F. A. A. Kingdom, and L. F. Gatley, "The standard deviation of luminance as a metric for contrast in random-dot images," *Perception* **19**, pp. 79–101, 1990.
2. J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of image," *IEEE Trans. Info. Theory* **20**, pp. 525–535, 1974.
3. F. Lukas and Z. Budrikis, "Picture Quality Prediction Based on a Visual Model," *IEEE Transactions on Communications* **30**(7), pp. 1679–1692, 1982.
4. N. Nill, "A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment," *IEEE Transactions on Communications* **33**(6), pp. 551–557, 1985.
5. S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, ed., pp. 179–206, 1993.
6. P. C. Teo and D. J. Heeger, "Perceptual image distortion," *Proc. SPIE* **2179**, pp. 127–141, 1994.
7. S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," *Intl. Conf. Acoustics, Speech, and Signal Processing* **4**, pp. 2351–2354, 1995.
8. J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models for Target Detection and Recognition*, E. Peli, ed., pp. 245–283, World Scientific, 1995.
9. C. J. van den Branden Lambrecht, "A working spatio-temporal model of the human visual system for image representation and quality assessment applications," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 2291–2294, May 1996.
10. Y. Lai and C. J. Kuo, "Image quality measurement using the haar wavelet," *Proc. SPIE: Wavelet Applications in Signal and Image Processing V* , 1997.
11. M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Transactions on Communications* **46**(9), pp. 1215–1226, 1998.
12. W. Osberger, N. Bergmann, and A. Maeder, "An automatic image quality assessment technique incorporating higher level perceptual factors," *Proc. IEEE Int. Conf. on Image Processing* **3**, pp. 414–418, 1998.
13. S. Winkler, "A perceptual distortion metric for digital color images," *Proc. IEEE Int. Conf. on Image Processing* **3**, pp. 399–403, 1998.

14. A. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Process.* **8**, pp. 717–730, May 1999.
15. S. Winkler, "Visual quality assessment using a contrast gain control model," in *IEEE Signal Processing Society Workshop on Multimedia Signal Processing*, pp. 527–532, September 1999.
16. J. Lubin et al., "Method and apparatus for assessing the visibility of differences between two image sequences," 1999. US Patent 5,974,159.
17. "Jndmetrix technology." http://www.sarnoff.com/. Sarnoff Corporation.
18. N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.* **9**, 2000.
19. M. Carnec, P. L. Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *ICIP 2003*, **2**, pp. 185–188, 2003.
20. http://foulard.ece.cornell.edu/VSNR.html.
21. Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, pp. 600–612, 2004.
22. H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing* **14**(12), pp. 2117–2128, 2005.
23. G. Zhai, W. Zhang, X. Yang, and Y. Xu, "Image quality assessment metrics based on multi-scale edge presentation," *IEEE Workshop on Signal Processing Systems Design and Implementation* , pp. 331–336, 2005.
24. A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Image Processing* **15**(2), pp. 422–429, 2006.
25. B. Girod, "What's wrong with mean-squared error?," in *Digital Images and Human Vision*, A. B. Watson, ed., pp. 207–220, 1993.
26. T. N. Pappas, T. A. Michel, and R. O. Hinds, "Supra-threshold perceptual image coding," *Proc. ICIP* , pp. 237–240, 1996.
27. W. Zeng, S. Daly, and S. Lei, "An overview of the visual optimization tools in jpeg 2000," *Signal Processing: Image Communication* **17**, pp. 85–104, 2001.
28. D. M. Chandler and S. S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," *IEEE Trans. Image Process.* **14**(4), pp. 397–410, 2005.
29. H. R. Sheikh, Z. Wang, A. C. Bovik, , and L. K. Cormack, "Image and video quality assessment research at live." Online. http://live.ece.utexas.edu/research/quality/.
30. R. L. DeValois and K. K. DeValois, *Spatial Vision*, Oxford University Press, 1990.
31. J. Schulkin, *Cognitive Adaptation*, Cambridge University Press, 1 ed., 2008.
32. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing* **15**(2), pp. 430–444, 2006.
33. M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, "A standard default color space for the internet–srgb." http://www.w3.org/Graphics/Color/sRGB, 1996.
34. S. S. Hemami, D. M. Chandler, B. C. Chern, and J. A. Moses, "Suprathreshold visual psychophysics and structure-based visual masking," in *Proc. Visual Communications and Image Processing 2006*, J. G. Apostolopoulos and A. Said, eds., **6077**, 2006.
35. http://vision.okstate.edu/csiq/.
36. D. M. Chandler and S. S. Hemami, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *J. Opt. Soc. Am. A* **20**, July 2003.
37. A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," *Pattern Recognition* **24**, pp. 1167–1186, May 1991.
38. E. C. Larson and D. M. Chandler, "Explaining crypsis and information content in the visual pathway using statistical properties of animal camouflage and natural scenes," *OSA Fall Vision Meeting* , 2007.
39. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**, pp. 2379–2394, 1987.
40. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," **37**, pp. 3311–3325, Vision Research 1997.

41. F. A. A. Kingdom, A. Hayes, and D. J. Field, "Sensitivity to contrast histogram differences in synthetic wavelet-textures," *Vis. Res.* **41**, pp. 585–598, 1995.
42. P. Kosevi, *Invariant Measures of Image Features from Phase Information.* PhD thesis, University of Western Australia, 1996.
43. D. M. Chandler and S. S. Hemami, "Suprathreshold image compression based on contrast allocation and global precedence," in *Proc. SPIE Human Vision and Electronic Imaging VIII*, B. E. Rogowitz and T. N. Pappas, eds., *Proceeding SPIE Human Vision and Electronic Imaging* , (Santa Clara, CA), 2003.
44. D. M. Chandler and S. S. Hemai, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Processing* **16**(9), pp. 2284–2298, 2007.
45. Z. W. H. R. Sheikh, A. C. Bovik, and L. K. Cormack. Image and Video Quality Assessment Research at LIVE [Online]. Available: http://live.ece.utexas.edu/research/quality/.
46. "Mad online supplement." http://lena.ceat.okstate.edu/MADsupplement. Oklahoma State University Image Coding and Analysis lab.
47. VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii," August 2003. http://www.vqeg.org.
48. H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," **15**, pp. 1349–1364, IEEE Transactions on Image Processing 2006.
49. A. K. Bera and C. M. Jarque, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Econ. Letters* **6**, pp. 255–259, 1980.