

Assessing Objective Image Quality Metrics for Bidirectional Texture Functions

Banafsheh Azari
CogVis/MMC,
Faculty of Media,
Bauhaus-University Weimar
Bauhausstrasse 11
99423 Weimar, Germany
banafsheh.azari@uni-
weimar.de

Sven Bertel
Usability,
Flensburg University of
Applied Sciences,
Kanzleistrasse 91-93,
24943 Flensburg, Germany
sven.bertel@hs-flensburg.de

Charles A. Wüthrich
CogVis/MMC,
Faculty of Media,
Bauhaus-University Weimar
Bauhausstrasse 11
99423 Weimar, Germany
charles.wuethrich@uni-
weimar.de

ABSTRACT

Bidirectional Texture Functions (BTFs) are view- and illumination-dependent textures used in rendering for accurate simulation of the complex reflectance behavior of fabrics. One major issue in BTF rendering is the large number and size of images which requires lots of storage. "Visually lossless" compression offers the potential to use higher compression levels without noticeable artifacts, but requires a rate-control strategy that adapts to image content and loss visibility.

In this contribution, we investigate the applicability of objective image quality metrics to predict levels of perception degradation for compressed BTF textures. We apply traditional error-sensitivity and structural similarity based approaches to predict levels of perceptibility for compressed BTF textures to achieve visually lossless compression. To confirm the validity of the present study, the results of an experimental study on how decreasing the BTF texture resolution influences the perceived quality of the rendered images with the results of the applied image quality metrics are compared.

In order to compare two representatives from each group were selected. The Visible Differences Predictor (VDP) and Visual Discrimination Model (VDM) are typical examples of an image quality metric based on error sensitivity, whereas the Structural SIMilarity index (SSIM) and Complex Wavelet Domain Structural Similarity Index (CWSSIM) are specific examples of a structural similarity quality measure.

Keywords

Perceptual experiment, Realistic rendering, Visual quality metric.

1 INTRODUCTION

To have a photo realistic display of fabrics, a real illumination and view dependent surface texture representation, called the Bidirectional Texture Function (BTF), was introduced in [1].

$$S_{\text{BTF}} = \int_{p \in \mathbf{P}} (\theta_i, \phi_i, x_p, y_p, \theta_o, \phi_o) \delta p, \quad (1)$$

BTF is a six-dimensional function representing the appearance of a material sample surface for variable illumination (θ_i, ϕ_i) and view (θ_o, ϕ_o) directions, where θ and ϕ are elevation and azimuthal angles, respectively, and (x, y) is the planar position over a material surface.

The three-dimensional textured models rendered through BTF rendering method are subject to various types of distortion during acquisition, synthesis, compression and processing. An appropriate image quality assessment scheme is a useful tool for evaluating image processing algorithms, specially algorithms designed to leave the image visually unchanged (e.g. compression algorithms) [2].

While the quality assessing task is simple for human observers, it actually involves very complex psychophysical mechanisms. Due to the high complexity of the human visual system (HVS), understanding it with current psychophysical knowledge is nearly impossible.

Currently, the only reliable way is to compare the overall visual similarity of two textures by independent observers in a psychophysical experiment [3–6]. However, this method is expensive, and it is usually too slow to be useful in real-world applications. As an alternative solution, BTF data modeling quality can be verified using objective image quality metrics (IQMs).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

This paper makes an attempt to validate these models with regard to predicting the visible quality differences in images rendered by compressed and non compressed BTFs.

For comparison of the traditional error-sensitivity and structural similarity based approaches, two representatives from each group were selected: The Visible Differences Predictor (VDP; [7]), Visual Discrimination Model (VDM; [8]), the Structural SIMilarity index (SSIM; [9]) and Complex Wavelet Domain Structural Similarity Index (CWSSIM; [10]).

The metrics were implemented and the results obtained from the predictions of the models were compared with each other and with the outcomes of a subjective quality measure experiment, which involved quality comparison tasks with pairs of textured objects of varying BTF quality levels [11].

In the next section, we will introduce the fundamentals of objective image quality assessment and review relevant full-reference objective quality metrics. Next some instances of the predictions of the models are presented and their performance is characterized accordingly. After discussion on the models, the detection results of metrics are compared with each other and with the outcomes of the user study, which is then followed by a conclusion and an outlook.

2 OBJECTIVE IMAGE QUALITY METRICS

The goal of Objective Quality Metrics is to design mathematical models that are able to predict the quality of an image accurately and automatically. An ideal method should be able to mimic the quality predictions of an average human observer.

Pixel-Based Metrics such as Root Mean Square (RMS) error or Peak Signal to Noise Ratios (PSNR) fail to assess the perceived degree of realism since they neglect important properties of the human visual system and poorly predict the differences between the images.

The philosophy used in constructing an objective image quality metrics is one of the major criterion employed for their classification. While traditional perceptual approaches to image quality assessment (bottom-up) are directly connected with the characteristics of HVS and try to simulate all the relevant components and psychophysical features as basic building blocks, and then combine them together, the ultimate goal of the structural similarity based approaches (top-down) is to make hypotheses about the overall functionality of the entire HVS and treat it as a black box, where only its input-output relationship is of concern. This section gives an overview of the general philosophy of both metrics and introduces the most popular and widely used metrics in each category.

2.1 Error Sensitivity Based Image Quality Measurement

A great variety of objective image quality assessment methods follow an error sensitivity based paradigm that attempts to analyze and quantify the error signal in a way simulating the characteristics of human visual error perception. In this part we will outline the perceptually driven image quality metrics that are used in this study that we will describe in the following, namely, VDP and VDM.

2.1.1 Visible Differences Predictor

The Visible Differences Predictor (VDP; [7]) is one of the well-known image distortion metrics, which consists of three main components: calibration of the input images, a HVS model and a method for displaying the visible differences.

The algorithm receives a pair of images (original and compressed images), and parameters for viewing conditions as input. After the calibration of the input images, in the next stage the HVS is modeled i.e. the lower-order processing of the visual system, such as the optics, retina, lateral geniculate nucleus, and striate cortex. The HVS model uses processes to limit the visual sensitivity.

Firstly, the original pixel intensities are compressed by the amplitude non-linearity based on the local luminance adaptation. Afterwards, the contrast sensitivity function (CSF) is processed to model the variations as a function of spatial frequency and so as to take into account the global state of luminance adaptation, orientation, image size and eccentricity from the fovea region. The sensitivity S as a function of ρ radial spatial frequency in c/deg is modeled by the following equation ([7]):

$$S(\rho, \theta, l, i^2, d, e) = P \cdot \min\left[S_1\left(\frac{P}{r_a \cdot r_e \cdot r_\theta}, l, i^2\right), S_1(\rho, l, i^2)\right], \quad (2)$$

where θ is the orientation in degrees, l is the light adaptation level in cm/m^2 , i^2 is the image size in visual degrees, d is lens accommodation due to distance in meter, and e is eccentricity in degrees. The parameters r_a , r_e and r_θ model the changes in resolution due to the accommodation level, eccentricity and orientation and P is the absolute peak sensitivity of the CSF.

The resulting images are decomposed into the spatial frequency and orientation channels using the cortex transform introduced by [12]. Cortex transform is a multi-resolution pyramid that simulates the spatial-frequency and orientation tuning of simple cells in the primary visual cortex. For every channel and every pixel, the global contrast and elevation of the detection threshold based on masking is calculated. This

detecting threshold is then used to normalize the contrast differences between target and mask images. The normalized differences are input into the psychometric function which estimates the probability of detection of differences for a given channel. This estimated probability value is summed across all channels for every pixel, and visualization of visible differences between the target and mask images is performed.

While this metric is designed for low dynamic range (LDR) images, [13] proposed an high dynamic range (HDR) extension of VDP, that can handle the full luminance range visible to the human eye. The modifications improve the prediction of perceivable differences in the full visible range of luminance and under the adaptation conditions corresponding to real scene observation.

2.1.2 Visual Discrimination Model

Another frequently used image discrimination measuring method is the Sarnoff Visual Discrimination Model (VDM; [8]). The Visual Discrimination Model acts in the spatial domain by firstly using an approximation of the point spread function of eye's optics, according to which the input data are convoluted. Next, the signals are re-sampled to be able to reproduce the sampling of photo-receptor in the retina. To break down the images into seven different resolutions, VDM uses a Laplacian pyramid [14]. At this stage each resolution must be one-half of the immediate higher image. Band-limited contrast computations are then performed.

In the next step, the selectivity of orientations in four different orientations is applied. To do this through steerable filters of Freeman and Adelson [15], a group of orientation filters were implemented. CSF was modelled through normalization of the output of every frequency-selective channel by the base-sensitivity for that channel. To implement masking, a nonlinear sigmoid is used. This is performed after convolving the errors at each level with disk-shaped kernels. Eventually, JND (Just Noticeable Differences) map or a distance measure is calculated as the Lp-norm of the responses of the masks. In the visual field of an observer, the eccentricity of images is an important factor. VDM is one of the few models that appropriately takes this into account. For color video, VDM was modified to the Sarnoff JND metric [8],

$$J = \frac{1}{\ln 2} \int_{V_{max}}^0 \sqrt{\frac{M(V)}{M_t(V)}} \frac{dV}{V}, \quad (3)$$

where V_{max} is the maximum spatial frequency displayed, $M(V)$ is the modulation transfer function of display and $M_t(V)$ is the threshold modulation transfer function of the human visual system.

2.2 Structural Distortion Based Image Quality Measurement

The fundamental principle of the structural approach is that the human visual system is highly adapted to extract structural information (the structure of objects) from the visual scene, and therefore a measurement of structural similarity (or distortion) should provide a good approximation of perceptual image quality.

In this section, we will mainly focus on two very recent and exceptionally successful general-purpose image quality assessment approaches, the Spatial Domain Structural Similarity Index (SSIM; [9]) approach and the Complex Wavelet Domain Structural Similarity Index (CWSSIM; [10]) approach. These approaches are based on high-level top-down hypotheses regarding the overall functionality of HVS (see [16]).

2.2.1 Spatial Domain Structural Similarity Index

Under the assumption that human visual perception is not built for detecting absolute, exact intensities, instead it is adapted to help us navigate the three-dimensional space we live in and, consequently, is highly adapted for extracting structural information from a scene, [9] introduced the Structural SIMilarity Index (SSIM).

In particular the SSIM index is a framework for quality assessment based on the degradation of structural information and is mostly sensitive to distortions that break down natural spatial correlation of an image such as blur, blocking, ringing, and noise.

The SSIM separates the task of measurement into three functions: Luminance $l(x, y)$, contrast $c(x, y)$, and structure $s(x, y)$. Given two images (or image patches) of x and y for comparison, the three similarity functions are then combined to yield the general form of the SSIM index structural similarity:

$$SSIM(x, y) = l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma, \quad (4)$$

where α, β, γ are positive constants used to weight each comparison function.

SSIM is a window-based algorithm that uses a square window, moving pixel-by-pixel over the image to measure loss of correlation, luminance distortion and contrast distortion locally [9]. To evaluate the overall image quality, a mean SSIM (MSSIM) index is calculated as follows:

$$MSSIM(X, Y) = \frac{1}{M} \sum_{i=1}^M SSIM(x_i, y_i), \quad (5)$$

where M is the number of samples in the quality map, x_i and y_i are the image contents at the i -th local window, and X, Y are the input images.

The structural similarity metric yields a result in a range of 0.0 to 1.0, where zero corresponds to a loss of all structural similarities and one corresponds to being an exact copy of the original image. Images with lighting-related distortions alone yield a high SSIM value while other distortions result in low similarities, corresponding well with the intuitive perception of quality.

2.2.2 Complex Wavelet Domain SSIM

A major drawback of the spatial domain SSIM algorithm is that it is highly sensitive to translation, scaling and rotation of images while perceptual metrics can successfully account for contrast and luminance masking, they are quite sensitive to spatial shifts, intensity shifts, contrast changes, and scale changes.

[10] suggested to implement a structural similarity metric in the complex wavelet domain and make it insensitive to these "non-structured" image distortions that are typically caused by the movement of image acquisition devices, rather than the changes in the structure of objects in the visual scene [10]. In addition, if an application requires an image quality metric that is unresponsive to spatial translation, this extension of SSIM can be adopted.

Given complex wavelet coefficients c_x and c_y that correspond to compared image patches x and y , the complex wavelet structural similarity (CWSSIM) is yielded by:

$$CWSSIM(c_x, c_y) = \frac{2 \left| \sum_{i=1}^N c_{x,i} c_{y,i}^* \right| + K}{\sum_{i=1}^N |c_{x,i}|^2 + |c_{y,i}|^2 + K}, \quad (6)$$

where c^* denotes the complex conjugate of c and K is a small positive constant.

The proposed algorithm shows some interesting connections with several computational models that have been successfully used to account for a variety of biological vision behaviors such as those pointed out by [17–19]. However, the algorithm does not provide any information on correspondences between the pixels of the two compared images and the method works only when the level of translation, scaling, and rotation is small (compared to the wavelet filter size).

3 EXPERIMENT

The goal of the experiment is to investigate the validation of error sensitivity and structural distortion based image quality metrics to predict the visible differences between compressed and non compressed texture resolutions. To achieve this, we analyze and compare the results of these models against each other and with the outcomes of the user study such as, performance of subjects (i.e., the subjects' ability to judge image quality differences) and the gaze data (locations and frequencies of fixations).

In the experimental study three datasets have been used. The first one was corduroy, available in the BTF database of the University Bonn¹, which we will refer to as *Cord-256*, whereby its texture pictures are 256x256 pixels. We generated two additional datasets by downscaling the *Cord-256* set through bilinear interpolation to respective resolutions of 128x128 pixels (*Cord-128*) and 64x64 pixels (*Cord-64*). For each of the three texture data sets, a three dimensional textured model of a sofa was rendered through the standard BTF rendering method at a screen resolution of 1920x1080 pixels.

The sofa model was oriented for presentation to the viewer so as to present textured parts across a large range of picture depth. We chose a sofa to have an object with a structured surface and composition (e.g., individual buttons, cushions, etc.). This is important in order to ensure that a large set of fitting BTF pictures will be selected as basis for the object's texture, with widely varying illumination and viewing angles (see Figure 1).

Pairs of images were displayed on a full screen in native resolution mode. Each pair consist of a sequentially presented rendering with the use of two of the three texture resolutions (256x256, 128x128 and 64x64). After the presentation of each pair, subjects were asked to make a decision about the comparative image quality within the pair: was the first or second image of better visual quality? Or were the two images of the same visual quality? A SMI RED250 remote eye tracking system was used in binocular mode with 250 Hz fixation detection, in order to record subjects' fixation behavior. A total number of 20 subjects, 12 males and 8 females, participated in the experiment, and they were not informed about the purpose of the experiment prior to conducting it.

The same sofa object model in the experimental study stimuli as well as one additional spherical object, which contains various angles and depth combinations, were utilized for making performance and detection results comparable with the outcomes of the experimental study. For the texture, two cases were considered including the Cord already known from the experimental study and Pulli, which is also available in the BTF database of the University Bonn.

Both objects are rendered in three levels of resolutions namely: 256x256, 128x128 and 64x64 pixels, which are referred to as *Cord-256 / Pulli-256*, *Cord-128 / Pulli-128* and *Cord-64 / Pulli-64*, respectively.

The images were presented on a 24-inch monitor with a resolution of 1920x1080 pixels at a distance of 70 cm from the viewer. The screen measured 22.35x15.80

¹ <http://btf.cs.uni-bonn.de/>.

	VDP Fixation location	VDM Fixation location	SSIM Fixation location	CWSSIM Fixation location
<i>Cord-256 _ Cord- 64</i>	-0.808	-0.1772	-0.498	-0.643
<i>Cord-128 _ Cord- 64</i>	-0.753	-0.1728	-0.320	-0.582
<i>Cord-256 _ Cord-128</i>	-0.015	-0.473	-0.155	-0.0617

Table 1: Correlations between IQMs results >75% of fixation location independently of presentation order. (p > 0.0001)

	VDP-Depth of the pixel	VDM-Depth of the pixel	SSIM-Depth of the pixel	CWSSIM-Depth of the pixel	Fixation-Depth of the pixel
<i>Cord-256 _ Cord- 64</i>	-0,1636	-0,6961	-0,0173	-0,2872	-0.2705
<i>Cord-128 _ Cord- 64</i>	-0,1124	-0,6929	-0,0092	-0,4498	-0.2305
<i>Cord-256 _ Cord-128</i>	-0,0061	-0,6413	-0,0055	-0,5105	-0.2405

Table 2: Correlations between IQMs results, number of fixation and depth of the pixel independently of presentation order. (p > 0.0001)

	#equal	#correct	VDM	SSIM	CWSSIM
<i>Cord-256 _ Cord- 64</i>	49	382	0.93624	0.963	0.822
<i>Cord-128 _ Cord- 64</i>	44	383	0.89378	0.971	0.838
<i>Cord-256 _ Cord-128</i>	423	63	0.39004	0.994	0.949

Table 3: Frequencies of correct answers, incorrect equal-quality answers (accumulated over all 20 subjects; sum of answers per pair: 480); and dprime value from VDM, SSIM and CWSSIM

inches and subtended approximately 33 degrees of visual angle. Due to the texture pattern, the minimal texture detail (i.e., for the parts of the sofa at the greatest depth in the image) had a cycle of 4 pixels, which means a subtended angle for a viewer of about 6 cycles per minute of a degree of arc. We employed the same condition for all the metrics.

3.1 Detection Results and Performances

In this section both, the output detection images of the image quality metrics, and the outcome of the user study are compared. The implemented metrics received pairs of images as input. The output detection images of the metrics were then compared and discussed.

For all the models, the following approach was employed [20]: the numerical value of the difference between images is the percentage of pixels for which the probability of difference detection is greater than 0.75. It is assumed, that the difference can be perceived for a given pixel when the probability value is greater than 0.75 (75%), which is the standard threshold value for discrimination tasks, [21]. This output value therefore ranges between 0 and 100, where 0 means the best result (no pixel with probability of difference detection greater than 0.75), while 100 means that all the pixel differences are above the difference detection threshold (the worst result).

However, since we also need a single overall quality measure, we use a mean index in the case of SSIM and CWSSIM models and JND for VDM. The index values fall within a range of 0 to 1, where 1 in JND

value of VDM means the worst quality, and 0 denotes an indistinguishable difference between the input images, which is in case of SSIM and CWSSIM mean index conversely.

Figures [1,2–4] present the output images of the metrics. To have a better comparison between metrics the results of two famous pixel-based metrics, the MSE and PSNR, for each image pair are also presented.

Next gaze fixation distributions of subjects across the sofa images were analyzed in order to assess whether differences exist for different image pair comparisons. Fixation counts for cells in an overlaid 16x16 grid are shown in Figure 1 (upper part) for three conditions.

Correlations between VDP/ VDM results (above 75%) and respective fixation location patterns can be observed in Table 1. We observed strong correlations between locations of predicted visually perceivable differences by VDP and observed fixation patterns only for Cord-256 and Cord-64 as well as Cord-128 and Cord-64, while significant, albeit a very poor correlation exists for VDM and fixation patterns for all image pairs. The results show a poor correlation for SSIM and CWSSIM.

In the next step, the responses of objective quality metrics to pixel depth for each image pair and the percentage of fixation in each depth were controlled.

Table 2 illustrates the correlation between IQMs responses and the depth of pixels as well as the correlation between fixation position and the depth of these pixels. The results show a poor correlation between



Figure 1: The output images of four IQMs by sofa with different 'Cord' texture resolution pairs. The color-scales on the right side indicate probability values of metrics in each pixel. The last row presents Just Noticeable Difference (JND) values of VDM, SSIM and CWSSIM. Additionally the MSE and PSNR, for each image pairs are also presented.

VDP, VDM and fixation and a significant correlation between SSIM, CWSSIM and pixel depth.

As shown by Figures 7, all curves react similarly to depth from quality perspective, but VDM is less sensitive than other metrics.

The first two columns of Table 3 illustrate the number of correct and equal answers yielded for each of image pairs, and the remaining columns present the result of VDM, SSIM and CWSSIM. The results show a significant correlation between subjects' ability to perceive differences between images and IQMs predictions.

In order to control the correlation between the saliency map (SM), Regions of Interest (ROI) and the responses

of IQMs, we followed [22] and computed a ROI map from the subjects' fixations.

The ROI map is a probability distribution of the gaze direction, therefore its integral is normalized to 1. Figure 5 (left-down) shows the ROI map obtained from individual fixations.

To define the saliency map the algorithm proposed by [23] was employed, with a new definition of the visual features (intensity, colour and orientation), which is the most popular in computer science, and has led to more convincing oculometric validations (see Figure 5 (right-up)). The Saliency Toolbox for Matlab, which is available online, was utilized in the present study [24].

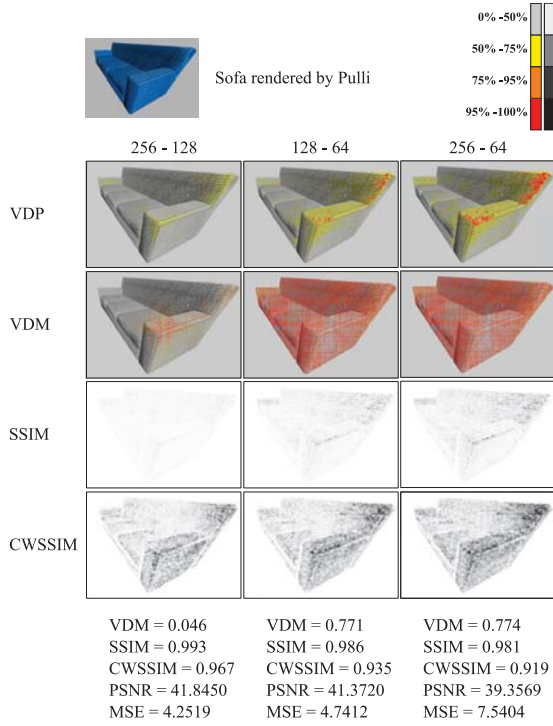


Figure 2: The output images of four IQMs by sofa with different 'Pulli' texture resolution.

	Total execution time (s)	
	Sofa	Sphere
VDP	7.256	4.654
VDM	0.340	0.152
SSIM	0.282	0.134
CWSSIM	0.929	0.432

Table 4: Total execution time in second. All the metrics run on the same machine.

Compared to the saliency maps shown in Figure 5, the ROI map is smoother. The saliency map and ROI are significantly correlated when $r = 0.560$ and $p < 0.001$.

Table 5 illustrates the correlation between IQM responses and ROI as well as the correlation between IQM responses and saliency map. As observed, the value between each IQM for ROI and saliency map is highly correlated when $r = 0.893$ and $p < 0.001$. The correlation coefficients between the adopted experimental subjective data set (ROI) and IQM responses exhibit that all models, except for VDM model, exhibit a good level of consistency with the subjective data.

The computation time is also another significant factor for selecting the image quality assessment. The computational complexity is measured in terms of time required by each of the metrics to assess the quality of a pair of images. In this step, each metric was computed for all pairs of images and then the average time was

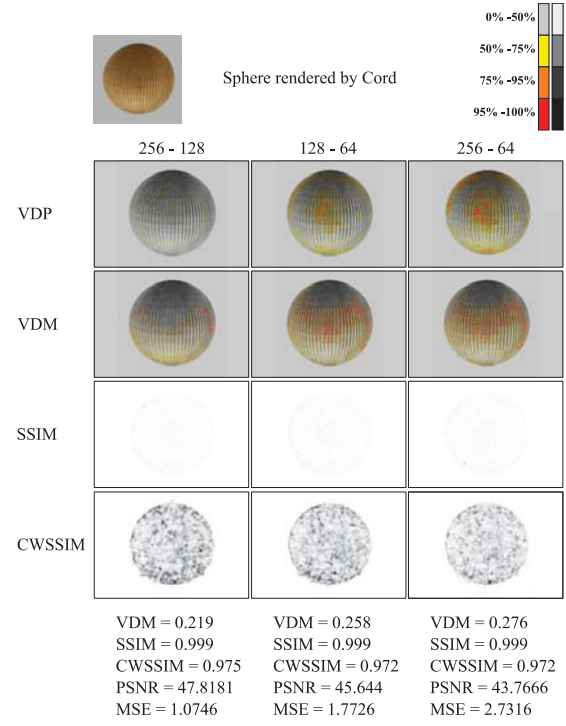


Figure 3: The output images of four IQMs by sphere with different 'Cord' texture resolution.

determined. The metrics were run on a computer with a 3.20 GHz Intel Xeon Six-Core processor.

In order to allow for a fair comparison, the publicly available Matlab implementation of each metric was used. The average performance of all the methods is provided in Table 4. SSIM, CWSSIM and VDM have a complexity of $O(N)$. This is due to the fact that these metrics work in the spatial domain avoiding the expensive FFT and FFT^{-1} transformations. This transformation can take up to 40% of the total execution time in VDP, and thus increase the complexity of this model to $O(N \log N)$ with an upper bound of $O(N^2)$ (see [25]). To control the reaction of the metrics to different geometrical distortions the object in the scene (sofa) was shifted without any other quality distortions and then used as a distorted image. Additionally we applied the metrics to blurred, salt & pepper and Gaussian noise contaminated images. Figure 6 illustrates the output detection images of the metrics.

4 DISCUSSION

The differences between the metrics are caused by placing pressure on different aspects of human visual perception. Nevertheless the results show that all metrics can be an appropriate replacement for subjective quality measurement matrices.

The vision models have different ways to visualize the detected probability. While VDP uses a psychometric

	VDP		VDM		SSIM		CWSSIM	
	ROI	SM	ROI	SM	ROI	SM	ROI	SM
<i>Cord-256 _ Cord- 64</i>	0.71	0.79	0.21	0.19	0.75	0.61	0.72	0.84
<i>Cord-128 _ Cord- 64</i>	0.63	0.57	0.23	0.20	0.73	0.58	0.83	0.85
<i>Cord-256 _ Cord- 128</i>	0.013	0.011	0.19	0.21	0.15	0.12	0.35	0.47

Table 5: Correlation between objective image quality metrics; VDP, VDM, SSIM and CWSSIM with ROI and saliency map ($p > 0.0001$).

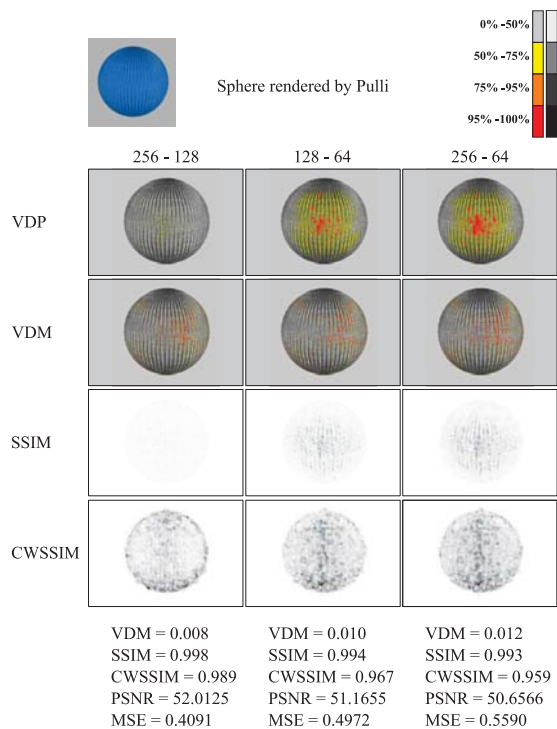


Figure 4: The output images of four IQMs by sphere with different 'Pulli' texture resolution.

function, which describes the relationship between the threshold contrasts and detection probabilities, to convert the normalized threshold contrasts into detection probabilities, all other models make direct use of JND map and neglect the psychometric function.

An advantage of the output map is that the nature of the difference can be observed and this observation can be used for further rendering optimizations.

The results show a significant correlation between subjects' ability to perceive existing differences between the images and predictions of VDM/CWSSIM models. Based on this investigation, it seems that VDM and CWSSIM can well predict the differences between two images.

The responses of objective quality metrics to pixel depth for each image pair shows that all react similarly to depth but VDM is less sensitive than other metrics. Due to the textured pattern, the texture details for the parts of the sofa from the depth of 0.3 to 0.8 have

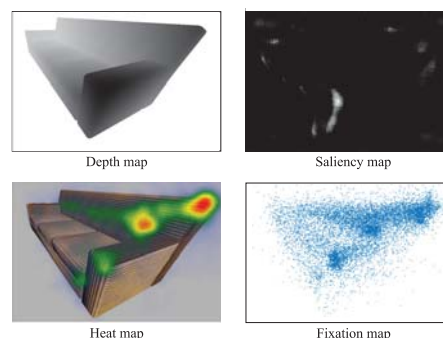


Figure 5: Depth map (left-up) ROI map (left-down), saliency map (right-up) and fixation map (right-down)

a 4 to 5 cycles per degree. HVS is most sensitive to intermediate ranges of spatial frequencies (around 4-6 cycles/degree), and is less sensitive to spatial frequencies both lower and higher than this. This explains why the metrics and the number of fixations have a higher rank in these depths.

The results of this experimental study showed that two groups of image comparisons exist. The first group consists of comparisons between Cord-256 and Cord-128. For this group, subjects are largely unable to perceive existing differences between the images. All models predict few visually perceivable differences for image pairs in this group.

The second group consists of comparisons between Cord-256 and Cord-64 as well as between Cord-128 and Cord-64. For this group, subjects are largely able to see the differences between the pairs. The models predict a larger number of differences which are also detectable with a higher probability.

Where strong correlations were observed between locations of predicted visually perceivable differences by VDP/CWSSIM and observed fixation patterns, a significant correlation was also observed between subjects' ability to perceive existing differences (the number of correct answers) and the results of VDM and CWSSIM tests (JND).

As observed, all models, except VDM, are able to detect regions of interest in images. This feature is promising for future research on ROI issues. The computation of VDM, SSIM and CWSSIM does not require

time consuming Fourier transformations (as VDP does) and they are certainly faster than that of VDP model.

Second, it was observed that all metrics are highly sensitive to small translations, scaling and rotations, which lead to high predicted perceptability values in metrics, even though no quality differences are available in compared images. In the frequency domain, small translations, rotations and scalings lead to consistent phase changes. Due to the fact that VDP works in frequency domains, it reacts with greater sensitivity to geometrical distortions than other metrics. According to [16], this problem can be overcome by analyzing images in complex wavelet domains through Structural Similarity based metrics, but the results were not promising in the case of the presented study.

Another common problem shared by the models is the disregard for color perception by HVS as well as incorporation of just the contrast sensitivity and luminance adaptation. A promising direction in the future would be an analysis of full-colored images.

Additionally, there is a lack of no-reference perceptual picture quality metrics, since both of the metrics are relative (full-reference). It is supposed that more work could be done in the field of no-reference image quality assessment.

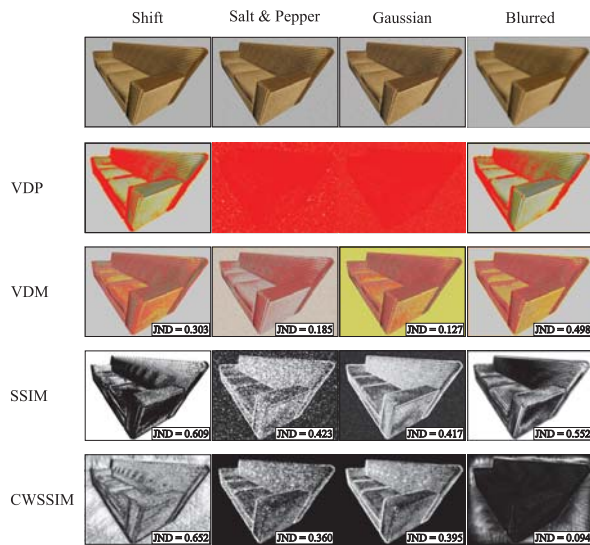


Figure 6: Objective quality metrics responses to shifted, salt & pepper and Gaussian noise contaminated and blurred images.

5 CONCLUSION

In the contribution, we investigated the suitability and integrity of certain image quality metrics, the traditional error-sensitivity and structural based to predict levels of perceptability for compressed BTF textures. To confirm the validity of obtained results, they were compared

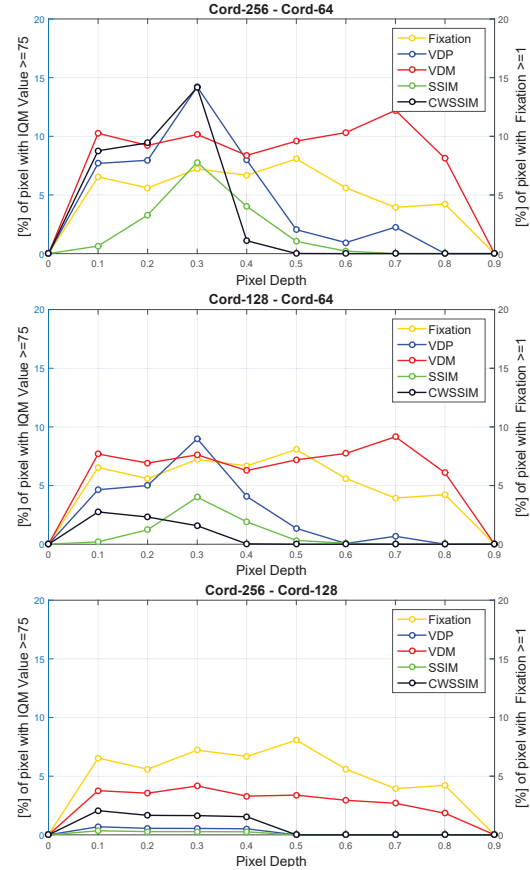


Figure 7: The percentage of fixation in each depth and the responses of VDP and VDM to the pixel depth between *Cord-256 _ Cord- 64* (top), *Cord-128 _ Cord- 64* (middle) and *Cord-256 _ Cord-128* (bottom).

with those obtained by an experimental study. In our validation experiment, it was observed that VDM and CWSSIM can in general better predict the differences between two images. On the other hand, VDP is better able to detect the location of visible differences in images.

Structural based IQMs are able to successfully predict image quality in close agreement with traditional error-sensitivity based IQMs.

The computation time is also another significant factor in image quality assessment, specially so when real-time image resolution changes need to be introduced as per the assessed quality of the rendered scene. In this scenario, all models, except VDP, prove to be proper options. This is because VDM, SSIM and CWSSIM operate in the spatial domain and unlike VDP, do not use the Fourier transform. However, in situations where one needs to improve the image quality of only parts of an object, only VDP can provide enough information on those areas requiring a higher resolution.

As observed, all models, except VDM, are able to detect regions of interest in images. This feature is promising for future research on ROI issues.

6 REFERENCES

- [1] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Transactions on Graphics (TOG)*, vol. 18, no. 1, pp. 1–34, 1999.
- [2] J. Filip and M. Haindl, "Bidirectional texture function modeling: A state of the art survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 11, pp. 1921–1940, 2009.
- [3] J. Meseth, G. Müller, R. Klein, F. Röder, and M. Arnold, "Verification of rendering quality from measured btfs," in *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pp. 127–134, ACM, 2006.
- [4] G. Müller, J. Meseth, and R. Klein, "Compression and real-time rendering of measured btfs using local pca," in *Vision, Modeling, and Visualization: Proceedings*, p. 271, AKA, 2003.
- [5] J. Filip, M. J. Chantler, and M. Haindl, "On optimal resampling of view and illumination dependent textures," in *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pp. 131–134, ACM, 2008.
- [6] J. Filip, M. J. Chantler, P. R. Green, and M. Haindl, "A psychophysically validated metric for bidirectional texture data reduction.," *ACM Trans. Graph.*, vol. 27, no. 5, p. 138, 2008.
- [7] S. Daly, "Digital images and human vision," ch. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, pp. 179–206, 1993.
- [8] J. Lubin, "A visual discrimination model for imaging system design and evaluation," *Vision models for target detection and recognition*, vol. 2, pp. 245–357, 1995.
- [9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [10] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 2, pp. ii–573, IEEE, 2005.
- [11] B. Azari, S. Bertel, and C. A. Wuethrich, "A perception-based threshold for bidirectional texture functions," *Proceedings of the 38th Annual Meeting of the Cognitive Science Society, CogSci 2016, Philadelphia, USA, August 10-13, 2016*, 2016.
- [12] A. B. Watson, "The cortex transform: rapid computation of simulated neural images," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 311–327, 1987.
- [13] R. Mantiuk, S. J. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images: model and its calibration," in *Electronic Imaging 2005*, pp. 204–214, International Society for Optics and Photonics, 2005.
- [14] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on communications*, vol. 31, no. 4, pp. 532–540, 1983.
- [15] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 13, no. 9, pp. 891–906, 1991.
- [16] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1–156, 2006.
- [17] J. A. Solomon and D. G. Pelli, "The visual filter mediating letter identification," *Nature*, vol. 369, no. 6479, pp. 395–397, 1994.
- [18] I. Ohzawa, G. C. DeAngelis, R. D. Freeman, et al., "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors," *Science*, vol. 249, no. 4972, pp. 1037–1041, 1990.
- [19] O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature neuroscience*, vol. 4, no. 8, p. 819, 2001.
- [20] K. Myszkowski, "The visible differences predictor: Applications to global illumination problems.," *Rendering Techniques*, vol. 98, pp. 223–236, 1998.
- [21] H. Wilson, "Psychophysical models of spatial vision and hyperacuity," *Vision and Visual Dysfunction*, vol. 10, pp. 179–206, 1991.
- [22] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 5, pp. 802–817, 2006.
- [23] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.
- [24] D. Walther, "Planes of the head." <http://www.saliencytoolbox.net/>, 2006.
- [25] B. Li, G. W. Meyer, and R. V. Klassen, "Comparison of two image quality models," in *Human Vision and Electronic Imaging III, San Jose, CA, USA, January 24, 1998*, pp. 98–109, 1998.