Steering Image Generation with Wavelet Based Perceptual Metric

Ajeetkumar Gaddipatti[1,2], Raghu Machiraju[3,4], Roni Yagel[1,5]

Department of Biomedical Engineering[1], Computer and Information Science[5], The Ohio State University
Department of Computer Science[3], NSF Engineering Research Center[4], Mississippi State University
Department of Biomedical Engineering[2], The Cleveland Clinic Foundation, Cleveland, U. S. A.

Abstract

It is often the case that images generated by image synthesis algorithms are judged by visual examination. The user resorts to an iterative refinement process of inspection and rendering until a satisfactory image is obtained. In this paper we propose quantitative metrics to compare images that arise from an image synthesis algorithm. The intent is to be able to guide the refinement process inherent in image synthesis. The Mean-Square-Error (MSE) has been traditionally employed to guide this process. However, it is not a viable metric for image synthesis control. We propose the use of a wavelet based perceptual metric which incorporates the frequency response of the Human Visual System. A useful aspect of the wavelet based metric is its ability to selectively measure the changes to structures of different sizes and scales in specific locations. Also, by resorting to the use of wavelets of various degrees of regularity, one can seek different levels of smoothness in an image. It is rare that such level of control can be obtained from a metric other than a wavelet based metric. We show the usefulness of our metric by examining it's effectiveness in providing insights for common operations of an image synthesis algorithm (e.g., blurring). We also provide some examples of it's use in rendering algorithms frequently used in graphics.

1. Introduction

The goal of any image synthesis algorithm is to produce a 2D image that is typically used for a variety of purposes. The image is judged for suitability by visual inspection and, if not deemed satisfactory a new image is generated. Suitability is often influenced by the rigors of the underlying application and the aesthetic leanings of the user. Seemingly appropriate parameters of the image synthesis algorithm are then altered to produce yet another image. This refinement loop is often continued a few number of times. It would be useful to employ a metric that guides the refinement process. The metric ideally should identify beneficial or adverse effects of various rendering operations on the final image and facilitate the choice of parameters. We propose a wavelet based perceptual metric in this paper. Objective perceptual metrics have been proposed elsewhere also [15]. The proposed metric differs from previous methods in that it has the ability to measure variations in images at specific locations, orientations and scales. Also, it has the ability to discern intensity functions of varying degree of smoothness in the image.

To explain the need for viable metrics let us consider the following examples derived from different image synthesis algorithms. For instance, in volumetric rendering blurring and aliasing can occur during the process of reconstruction along the ray (raycasting algorithms [9]) or along a slice (slicing algorithms [23]). Thus, if a certain filter of a certain size is used in function reconstruction, we would like to gauge the effect of this filter at different locations in the final image and on structures of different sizes. Also, more importantly, we would like to determine if a more pleasing and useful image can be obtained by altering the size of the reconstruction filter or even replacing the existing filter with another one. Through successive comparison of images we

should ideally choose the correct size of the reconstruction filter or choose one filter over the other. Similarly, in both surface and volumetric raytracing algorithms, to mitigate the effects of aliasing, sub-pixel rays are cast as part of a supersampling scheme. The subpixels are then filtered to result in a pixel intensity. Once again, it would be useful to employ a quantitative tool to determine if supersampling actually improved the image quality and to determine the size of the subpixel neighborhood. Also, it might be useful to understand the efficacy of jittering schemes which are commonly used to anti-alias raytraced images. Yet another case for the use of numerical metrics can be made when classification and shading techniques are employed to enhance structures. Noise is often introduced in the image as a result of these enhancement techniques. Once again quantitative insight is often useful. Although our discussion pertains to rendering algorithms, a similar case can be made for other image synthesis algorithms including radiosity.

The mean square error (*MSE*) and the root mean square error (R*MSE*) are often used to compare images. However, they are not adequate measures to employ in comparing images. We shall provide several examples of the inadequacy of *MSE* in Section 5. The limitations of *MSE* arise from the fact that comparisons are based only on corresponding pixels and does not include any knowledge of the human visual system (*HVS*) and the underlying structures in the image. What we propose here are numerical metrics that allow control of the refinement process of an image synthesis algorithm. Ideally, one would seek to devise a metric that

- measures the information content of an image in terms of structures and features,
- responds to changes wrought to an image from the application of operators, such as blurring, and
- incorporates aspects of the *HVS*.

Our goals are similar in essence to those of Higgins who states that the image quality of tone reproduction systems depends on how well the microscopic details are reproduced [6]. We propose a perceptual metric that meets these three goals. Each of the two images to be compared is subject to a process to yield saliency values which are then actually compared in a *MSE* fashion. The saliency value of an image is determined by first conducting an orthogonal wavelet transform. A simple structure detection scheme follows that allows identification of all significant scale-coherent structures (edges, corners, ridges, etc.) in an image. Finally, the coefficients of the wavelet transform are modulated with a *Contrast Sensitivity Function* (*CSF*). The *CSF* measures the frequency response of the *HVS*. Essentially, the proposed metric is a weighted *MSE*, the weights being derived from the *CSF*.

In Section 2 we present the necessary background on wavelet transforms, and make the case that one can obtain metrics (based on the wavelet transform) which emulate the functioning of the human visual system. Previous work in the area of image comparison metrics is described in Section 3. In Section 4 we describe our metric, while in Section 5 we provide some examples of possible usage. Finally, in Section 6 we provide concluding remarks and point to the future.

2. Multiscale Approach to Perception

Multiscale transforms have been employed to analyze and code images. We examine their suitability to emulate the human visual system. Then, we justify our choice of the wavelet transform for developing perceptual metric. Finally, in this section we describe some previous efforts from various application domains.

Of the many multiscale transforms used in image analysis, we are particularly interested in the *Gabor* [14], the *Cortical* [18] and the wavelet transforms [3]. The first two transformations are more amenable to use as tools for image analysis rather than coding or compression of images. It has also been shown that the receptive field of the visual cortex can be modeled as a 2D Gabor Transform [14]. For 2D images, the Gabor transform also describes the orientation of the structures. However, the transform is redundant and is expensive to compute since it cannot be realized as a pyramid. There do exist bi-orthogonal approximations to the Gabor Transformation. Notable among these is the cortical transformation of Watson [18] which has often served as a computational model of the *HVS*. The cortical transformation was devised as a result of a study of the human cortex and the visual system. It differs from the Gabor transform in two main ways; the basis functions are Gabor-like (but are different) and the transform can be realized as a pyramid. Although the motivation for the multi-resolution nature was more for the sake of efficiency, it was noted by many that the *HVS* does function in a multi-scale fashion. Witkin [22], for instance, noted that the eye discerns only those features that persist through all scales.

The discrete orthogonal wavelet transform does not serve as a computational model for the *HVS*. However, it can be computed very efficiently and can represent arbitrary functions with few number of coefficients. The basic idea behind wavelet analysis is that the original image is split into several sub-images by a dyadic subdivision of image's frequency spectrum. The process is similar to that of *HVS* wherein image is locally decomposed into spatial frequency and orientation components. Finally, by choosing wavelets with different degrees of regularity one can choose the level of smoothness one is seeking in the intensity function. We employ Daubechies wavelets [3] in the work described here. These wavelets are indexed by a parameter N - a ψ_D^N function is implemented as a $2N$ tap band-pass filter. Two important properties of these filters are that they have compact support in the spatial domain and that they are N-regular (ψ has N vanishing moments), i.e., all polynomials of degree $p \le N$ are accurately represented, as shown below.

$$\int x^p \psi_D^N(x) \, dx = 0 \qquad 0 \le p \le N \tag{1}$$

There do exist disadvantages over the use of the orthogonal wavelet transform, namely that the basis functions that describe a local region of an image severely alias each other. It is only in the inverse transform that the aliasing effects are mitigated. Redundant transforms like the steerable transforms found in [16] perform a decomposition into frequency bands which are further split into orientation bands. In redundant transforms, basis functions overlap each other at different scales, thus reducing the aliasing artifact. However, we do not explore the use of this transform in this paper.

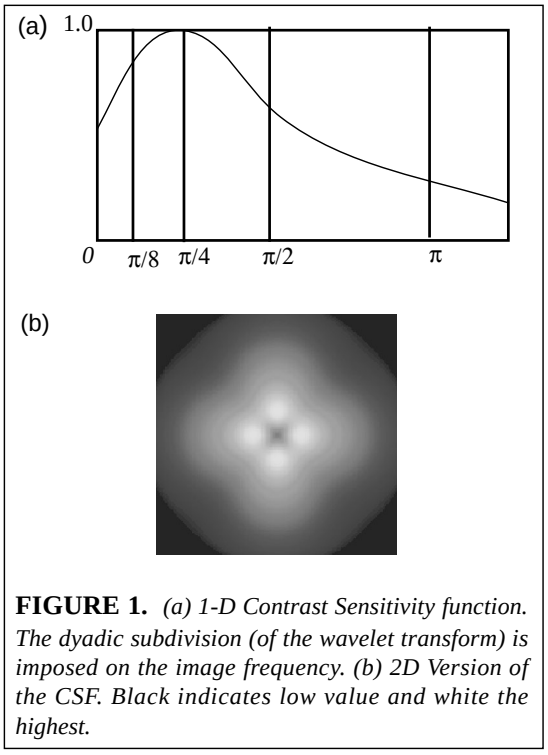

FIGURE 1. *(a) 1-D Contrast Sensitivity function. The dyadic subdivision (of the wavelet transform) is imposed on the image frequency. (b) 2D Version of the CSF. Black indicates low value and white the highest.*

We now describe a popular *CSF*. Later, in Section 4, we shall describe a way to employ the contrast sensitivity function with a pyramid transform. The human cortex is often modeled as a linear system and its response to a visual excitation in the receptive field is expressed as a convolution of the impulse response of the visual cortex with the input stimulus. Since convolution over a large array of neurons in the cortex is expensive to compute, the impulse response is often expressed as a *Contrast Sensitivity Function*. Thus, the response to an input can be obtained by multiplying the *CSF* with the Fourier transform of the input. The *CSF* measures the response of the visual system to different frequencies. Another perspective of *CSF* is that it is the reciprocal of the contrast necessary for a given frequency to be perceived. It is important to note that the response is lower

for higher frequencies. There exists reported work which attempts to measure the *CSF* of the human visual system. For instance, Mannos and Sakrison [10], after conducting a series of psychophysical experiments on human subjects, found that the *CSF* can be modeled by the function in Equation 2.

$$C_s(f) = [0.0499 + 0.2964 f_s] \times \exp\left[-(0.114 f_s)^{1.1}\right] \tag{2}$$

Here f_s is spatial frequency in cycles per degree. This model converted to cycles per pixel (*cpp*) using the following transformation [8]:

$$f_s\left(\frac{cycles}{degree}\right) = f_i\left(\frac{cycles}{pixel}\right) \times f_n\left(\frac{pixels}{degree}\right) \tag{3}$$

$$= \frac{f_i}{arc\sin\left\{\frac{1}{\sqrt{\left[1 + d^2\left(h^2 + v^2\right)\right]}}\right\}}$$

Here, f_i is the image frequency (obtained from Fourier transform), the normalizing factor f_n is the number of pixels within *1* degree at the viewing distance of *d* times the diagonal image size, *h* is the horizontal image size and *v* is the vertical image size in pixels. A 2D version can be easily obtained by replacing f_i with radial frequency, $f_r = \sqrt{\mu^2 + \upsilon^2}$, where μ and υ are the horizontal and vertical frequencies.

Figure 1(b) shows the 2D contrast sensitivity function. Magnitude of the function decays radially. White indicates high value and black low. Other *CSF* functions have also appeared in the literature. Some of these have been described for photographic and tone reproduction systems [6] and used to describe the resulting image quality. We, however, shall restrict ourselves to the one constructed by Mannos and Sakrison [10].

One can conceivably multiply the *CSF* with the frequency spectrum of an image and then determine a saliency [15]. However, this disallows the ability to delineate regions of interest and the use of simple schemes to detect structures. Hence, multiscale transforms offer significant advantages when combined with a *CSF.* As an alternative, a windowed Fourier Transform can be conducted to maintain locality [11] [21]. This methodology is however computationally expensive. We propose use of an alternate methodology which determines the weight of the *CSF* curve in each subband. A frequency decomposition induced by a dyadic wavelet transform is imposed on the *CSF* (Figure 1a). Thus for each band a weight can be computed which is then applied to the wavelet transform. The *CSF* based model is linear and cannot explain all aspects of human vision, especially the ability to detect low contrast images. In [17] the transform of [16] is used to construct a non-linear model of human vision which can address the issue of contrast masking. We however, limit ourselves to the linear perceptual model. We provide the details of this process in Section 4. We now describe previous work in designing metrics for variety of applications including image synthesis, coding and database searches.

3. Previous Work

It should be noted that we are primarily interested in objective measures and do not survey the vast literature that exists on subjective measures of image quality. We are cognizant of the fact that our objective measures have to be validated by an evaluation procedure that employs human subjects. However, objective measures can perform well in many situations if designed with care. For instance, it was reported in [2] that the quantitative measure of signal-to-noise ratio (SNR) actually outperformed more expensive subjective tests.

There has been little effort in developing viable image metrics in image synthesis. Among the first known efforts, Rushmeier *et al.* employ metrics from image coding literature to compare images that arise from radiosity algorithms [15]. Three metrics were described which differ from each other in the *CSF*s used. Five evaluation criteria were proposed to determine the suitability of the metrics. Sample scenes were constructed and the behavior of the metrics was examined for these scenes. Although two of the proposed metrics (including

the one in [10]) performed adequately for the given scenes, they still did not capture local variations in intensity well. The metrics employ the Fourier transform which does not preserve local spatial characteristics. Finally, the metrics were designed and used more for purposes of visual similitude rather than as guiding tools in a refinement process. In [1] methodologies have been proposed to include perceptual considerations into the rendering process. Although laudable, these methodologies are cumbersome to incorporate and are restricted to specific rendering algorithms.

Much effort has been expended in incorporating the characteristics of the *HVS* into all aspects of image coding and restoration [8][19][20]. For instance, the *HVS* weighing method of Mannos and Sakrison [10] has been applied to coders including those based on the wavelet transform. However, as it was pointed out in [15], one cannot directly use the metrics of coding in image synthesis. In medical imaging, however, much effort has been expended in defining metrics. Furuie *et al.* designed a set of phantom images which test the reconstruction accuracy of algorithms on 3D PET (Positron Emission Tomography) images [5]. As a result, they design Figure-Of-Merits (FOMs) which can be used to benchmark algorithms. Also, Cosman *et al.* performed a study of quality measures which are applied to compressed medical images [2].

Similarly, considerable work has been done to develop viable metrics for searching in image databases. These metrics operate on color, texture, orientation, shape and contrast of images. Essentially, one searches for similitude in image content in an extracted feature space.The Haar wavelet transform was employed in [7]. However the resulting metric is not robust. Finally, in multispectral image fusion, images obtained from different spectral bands are combined [21]. The combination is guided by the presence of structures or information. Wavelet based methods are being reported which estimate the features or structures in an image region-by-region. The maximum estimator, the energy, or a *HVS* weighted measure is used to determine the closeness of two images. The images are combined if the images are close, otherwise the image with the larger estimate dominates. Wilson *et al.* also employ the *HVS* weighing method of Mannos and Sakrison [10] to guide the combining process. A windowed Fourier transform of the wavelet sub-bands is performed to obtain a finer frequency analysis of the wavelet spectrum. These frequencies are weighed by the contrast sensitivity function. We proposed a metric for image synthesis [11] similar in essence. However, for reasons of computational efficiency we propose another metric which is easier to compute. We describe our methods in Section 4.

4. Wavelet Based Image Comparison Metric

In this section we describe our metric. The perceptual metric we propose here builds on the subband coherent structure detection algorithm described in [12]**.** The detection of coherent structure follows the wavelet transform. Later, by employing weights based on the *CSF* we modulate the wavelet coefficients. Finally, the images are compared in the mean square sense.

4.1 Coherent Structure Detection

Our sub-band combining algorithm described in [12] is similar to the one presented in Xu *et al.* [23]. Both attempt to detect correlated structures across the scales by employing the product of the wavelet transform across scales. However, in [23] the emphasis is on denoising and being able to detect edges, while in [12] the algorithm identifies structures of arbitrary singularity and assigns pixel-wise saliency values in regions populated by singular structures. The saliency is obtained from a mask derived from the process of multiplication. Figure 2 shows examples of images and their masks. The mask persists in regions of discontinuities and is zero otherwise. Wherever the mask is set to zero, the corresponding coefficients in the pyramid are set to zero. This mask can be used to further threshold the coefficients. Denoising is also implemented in two stages of the algorithm. The coefficients are first thresholded by using a Gaussian noise model as reported in [4]. Later, while generating the combining mask, denoising is again achieved as the noise in smoother levels is eliminated. As a result of the second thresholding operation, only the coefficients contributing to coherent strutures remain.

4.2 Perceptive Modulation

The coefficients at each level and location are modulated using a perceptual weight arrived at by integrating *CSF* over the area occupied by the band in 2D frequency space. We use the *CSF* model in Equation 2, substituting radial frequency f_r for f_i. The wavelet transform is obtained from the continuous application of two quadrature mirror filters [13]; *H*, a low-pass filter and *G*, a half band-pass filter. Thus, by applying *G* to a discrete input with bandwidth $(0,\pi)$, a signal with bandwidth $(0,\pi/2)$ is acquired. Thus, after *m* iterations the

weight for level *m* is:

$$C_m = \frac{\int_{FB_m} CSF\omega d\omega}{\int_{FB_m} d\omega} \quad \text{where} \quad FB_m = \left(\frac{\pi}{2^m} , \frac{\pi}{2^{m-1}} \right) \tag{4}$$

This weight measures the area of the *CSF* curve in each octave band (Figure 1a). It is assumed that the filters *H* and *G* are ideal and have non-aliased spectra. For *2D* frequency domain, the integral is over a rectangular area rather than a line. The weights now measure volume under CSF surface. Each level *m* has three weights except the coarsest level which has a weight only for approximation. The weights are precomputed and stored in a table.

4.3 Perceptual Metric

The perceptual comparison metric is computed as

$$M_p(A, B) = \frac{\sum_{m,i,j} \left| S_A(m, i, j) - S_B(m, i, j) \right|^2}{N_h \times N_v} \tag{5}$$

where *A* and *B* are the images being compared and S_A and S_B are saliency values of pixel *i, j,* at level *m* of the wavelet pyramids. N_h and N_v are number of pixels in horizontal and vertical directions. Saliency measures the response of the *HVS* to the frequency input at pixel (*i,j*) in the m^{th} sub-band image. Thus,

$$S_A(m, i, j) = C_m W(A, m, i, j) \tag{6}$$

Here C_m is the contrast sensitivity weight derived in Equation 4 and *W*(*m,i,j*) is the wavelet coefficient at level *m* and pixel location (*i,j*). Using saliency values at low resolution levels only gives us low resolution distortion metric. Similarly, vertical distortion can be obtained by considering saliency values in vertical orientation only. Thus, one has the ability to measure changes or distortions in sections of the image and along different orientations (vertical, horizontal and diagonal). One can also compute a composite distortion over all levels and orientations, which is essentially the sum of saliencies of all blocks at all levels. We now provide experiments to illustrate the efficacy of our metric. We specifically will show that incorporating perceptual criterion is critical to the development of metrics for image synthesis.

5. Experiments

In this section we describe the response of our perceptual metric to different image processing and rendering algorithms. First, we show why the *MSE* is a not a viable metric. Later, we study the responses of the *MSE* and the perceptual metric to blurring, supersampling and an example image synthesis algorithm. We start with an image *A* and create the set $Q = \{A_1, \ldots, A_N\}$. We now measure the distance of consecutive images in terms of the total metric or in terms of sub-bands or limit ourselves to regions. Effectively, for a metric *M*, we measure the incremental distortion set, $\{M(A, A_1), \ldots, M(A_{N-1}, A_N)\}$. This set of measurements emulates the refinement loop of an image synthesis algorithm.

We use a Daubechies wavelet of length *8* to perform *4*-level wavelet transformation of all images. The choice of a particular Daubechies wavelet is dictated by the inherent smoothness of the image.

It is generally the case that Daubechies wavelets with eight coefficients suffice for most images. In the case of color images we convert the *RGB* images to the *L*a*b* color space and perform the wavelet transform on each of the components of this space.

Before we describe our experimental methodology we enumerate a list of desirable properties an image com-

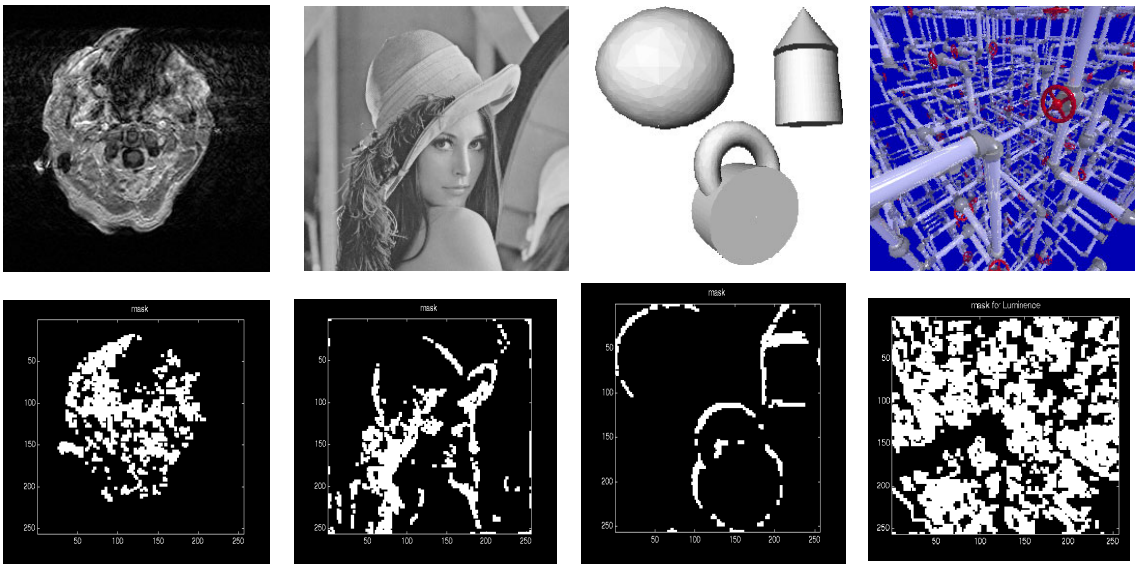

FIGURE 2. *Images and their masks. The top row shows the images of size 256^2, while the bottom row shows their masks. The mask is obtained with a Daubechies wavelet of length 8. The mask is set wherever a discontinuity exists, (a) MRI (b) Lenna (c) Synthetic (d) Pipes.*

parison metric should possess. Any metric of image comparison should measure numerical distortion and visual similitude. If two images *A*, *B* were to be compared, we expect the metric *M* to behave in the following fashion:

- *M(A,B)* = 0 if *A*, *B* have identical features
- *M(A,B)* is small if *A*, *B* have some dissimilar features
- *M(A,B)* is large if *A*, *B* are very dissimilar.

These set of rules dictate that the metric should not behave in a monotonic manner in response to changes in the applied operator. The behavior should rather be dictated by the changes wrought upon the image. Rushmeier *et al.* used few ratio based criteria in addition to the ones listed above [15]. We did not consider ratios in our metric as it is already normalized. Before we describe our experimental methodology, we illustrate the short comings of the *MSE* metric.

5.1 Why Is MSE Not A Viable Metric?

We illustrate this through the example shown in Figure 3. The photograph of Albert Einstein is distorted in two different ways. In Figure 3b, a constant bias is added, while in Figure 3c the edges of the image are altered resulting in greater visual distortion. The photographs are texture mapped on to a polygonal model of an ashtray and rendered with a raycasting algorithm. The distorted images are then compared with the original using *MSE* and our metric. The *MSE* measured at *225* for the image in Figure 3b and *160* for the image in Figure 3c. This indicates that the image in Figure 3c is a better image as it is closer to the original. However, this certainly is not the case. The perceptual metric on the other hand provides a realistic estimate of the image quality. Our metric measured at *156* for the more distorted image in Figure 3c and *73* for the less distorted one in Figure 3b.

We now summarize the reasons for *MSE* not being a viable metric in rendering algorithms:

- An important reason why *MSE* fails to report changes in an image is that it is a pixel based measure.In

Figure 3b the visual distortion is not significant since a constant bias (d.c value) is added and the human visual system does not notice extremely low frequency information. However, the pixel-wise distortion is maximum in this case. While in Figure 3c, the visual distortion is significant since edges are altered considerably resulting in a larger amount of high frequency information which the *HVS* detects. However, the pixel-wise distortion is smaller since fewer pixels are altered.

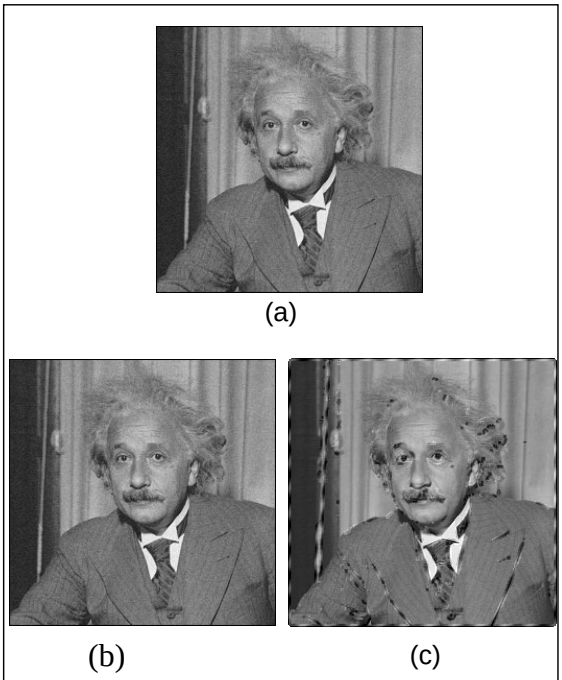

FIGURE 3. *Example of why MSE is not viable metric. (a) Original (b) Image with a constant bias (c) Image with distorted edges. The MSE between the original and distorted images leads one to believe that image (b) is more distorted. In reality, image (c) is more distorted.*

• Another major characteristic of *MSE* is that it always depicts a monotonic behavior when presented with incremental, degraded/enhanced images. This is not always true in many rendering situations, especially when the effects of various operators are nonlinear. This behavior of *MSE* violates the three conditions we laid out for a metric in the beginning of this section. We now explore the behavior of the metric on commonly used operator namely, blurring.

5.2 Smoothing

We chose three representative images to study the effect of the blurring operator.

• *MRI* (Figure 2a)- This image contains sufficient noise and has a profusion of edges and small structures. This image is typical of images arising from rendering of medical datasets.
• *Lenna* (Figure 2b) - This image has significant amount of texture (hair) and smoother edges.
• *Synthetic (*Figure 2c*)*- This image has several sharp edges and is representative of images used in computer graphics.

Although many filters can be employed the Gaussian filter is commonly used to perform smoothing. We show the effect of using increasingly larger box filters over an image of size 256^2. This is shown for all three representative images. Each of the images are subjected to smoothing by a Gaussian filter of size

$2m + 1, m = 1, \ldots, 4$. Blurring occurs from the application of a low pass filter which essentially removes high frequency content in an image. Larger size filters decimate the frequency content more significantly. The perceptual metric (Figure 4a) for the *Synthetic* image drops constantly, similar in behavior to the corresponding plot for *MSE* (Figure 4b). However, for the *MRI* and the *Lenna* images, while the *MSE* monotonically drops in value, the perceptual metric rises first and then drops. This indicates that most significant change occurs between the second and third images, rather than the first and second as the *MSE* indicates. A visual examination of the images (Figure 4c) confirms this fact. A possible explanation could be the fact that the blurring operator of size 7 (*=2*3+1*) alters many more structures significantly. In the case of the *Synthetic* image, the structures are much larger in pixel area and hence no such behavior is observed. In the case of all three images the difference in the metric is smaller as the filter gets larger in extent. It is for smaller filter sizes that the change is more dramatic. This is evident from a visual inspection of the images also. Finally, the *MSE* shows less degradation for the *MRI* image than the *Synthetic* image which is not the case. This can be attributed to the fact that the *MRI* image is noisy and populated with small features and hence the pixel-wise differences show up prominently. Our implementation of Rushmeier metric[15] followed *MSE* closely for this example. In the following examples also, Rushmeier metric fared worser when compared to our metric. Due to space constraints, we do not discuss this any further in this paper.

Another benefit of the wavelet based perceptual metric is the ability to observe the effect of operators at different scales (Figure 4d). More importantly, the effect on structures of different sizes can be studied. For example, it is instructional to see that most of the higher frequency (*k=1*) difference spectra is diminished rather quickly. At this level the details arise mostly from noise and the difference in the saliencies is very insignificant. For the next level (*k=2*), the change in the metric is less dramatic, although still significant. It is the spectra at level *k=3*, which dominates the total spectrum and resembles the Figure 4a closely. The approximation spectrum hardly changes. This is expected since blurring should only affect higher frequency components. It is these insights which make the wavelet based perceptual metric useful. It is often the case that a different smoothing or low-pass filter provides a better image. The difference at times cannot be easily perceived. It is possible to see improvement if any, with the aid of the perceptual metric, once again.

5.3 Aliasing in RayTracing

To illustrate the effects of aliasing, we employ supersampling schemes which gradually decrease the high frequency content in an image. In a typical supersampling scheme the benefits are accrued when the higher frequencies of the primary spectra are not aliased with those from neighboring spectra. Supersampling increases the width between the spectra reducing the chance of an overlap and hence aliasing. We used subsampling and supersampling schemes on the *Pipes* image (Figure 2a) to illustrate efficacy of the metric. The subsampling was done by taking every 4^{th} (image 1), 2^{nd} (image 2) pixel. The third image resulted from shooting a single ray from every pixel. The other images were generated by using supersampling and sub-pixel neighborhoods of 2^2,...,4^2. The aliasing is reduced and the images, which are not shown here, look significantly better. The *MSE* metric once again shows a monotonic behavior and indicates a much more significant improvement than is achieved in reality. The perceptual metric on the other hand indicates more modest changes. Both metrics are normalized with respect to number of pixels in the image. The perceptual metric indicates that a supersampling of size *3* is sufficient, which corresponds to the result of visual examination.

We also conducted the same test on the *PacMan* image shown in Figure 5b. Only a supersampling scheme was conducted for this image beginning with a neighborhood of *1x1*. This image has texture and structures of different sizes. Figure 5c shows the plot of *MSE* and our metric for unjittered image sequence. Figure 5d is a plot for jittered image sequence. As expected, the values for the *MSE* and perceptual metric were more for the jittered image sequence.

5.4 Artifacts in Volume Rendering

We tested our metric on images obtained from a slice-based volume rendering algorithm for unstructured grids generated from tetrahedral cells [24]. The algorithm slices the unstructured grid parallel to the screen and then composites the resulting triangles from each slice. The algorithm employs an edge-table scheme to exploit the coherency in such grids. The quality of the image depends on the number of slices. The images suffer primarily from aliasing. We employ our method to determine the appropriate number of slices. Once again, we find the *MSE* to be inadequate. It seems to indicate that the third image is acceptable enough. However, a visual display of the images in Figure 6 indicates otherwise. Real improvements start showing up in the sixth image (*60* slices).

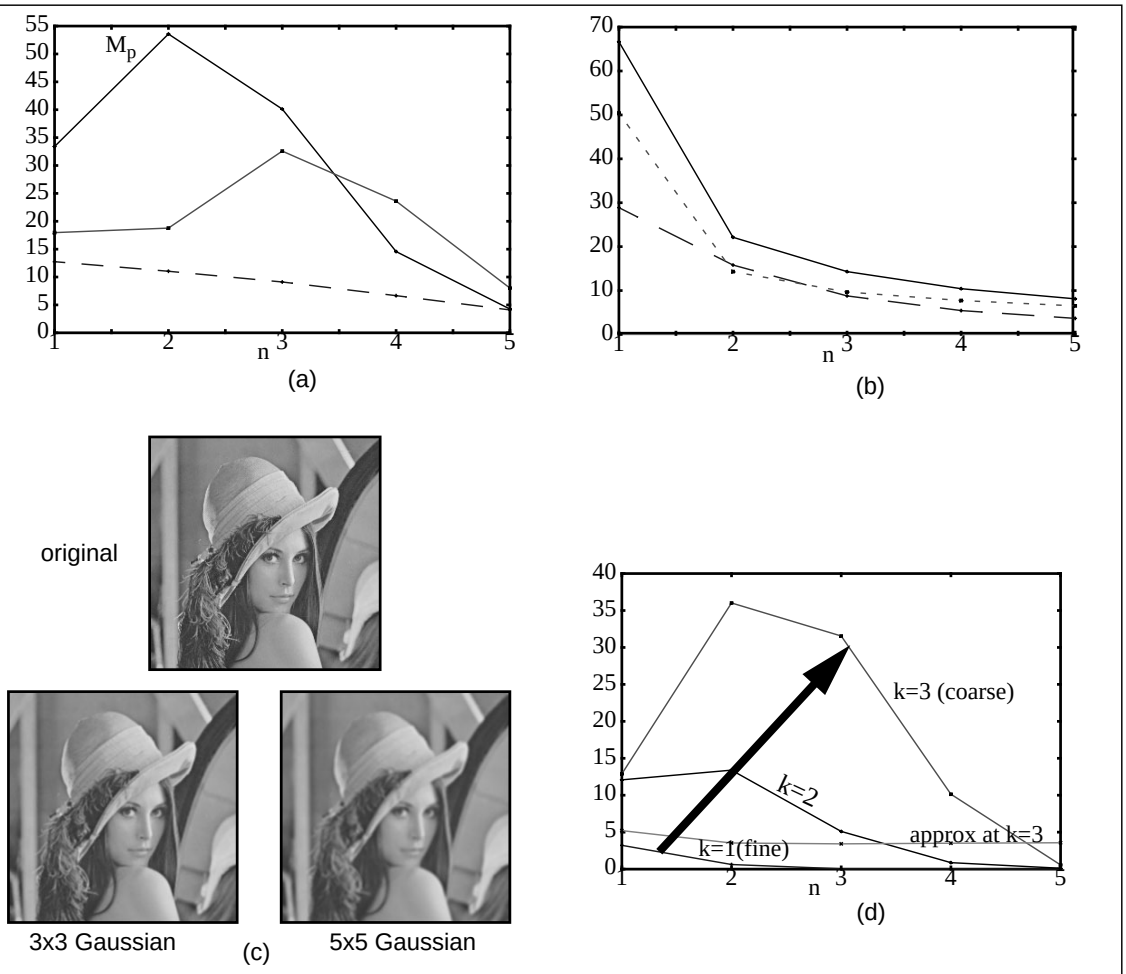

FIGURE 4. *Effect of Gaussian Blurring. We show the effect of the Gaussian filter on three images (MRI - dashed line, Lenna - solid line, Synthetic - dotted line). The x-axis indicates the successive image pairs compared (n), while the vertical axis measures the metric. (a) Incremental behavior of perceptual metric, M_p. (b) Incremental behavior of MSE, M_s.(c) The sequence of first three Lenna images obtained by applying a 3x3 and 5x5 Gaussian filter respectively. It is important to note that maximum distortion occurs between the second and third image. (d) Incremental scale distortions for Lenna image. The arrow indicates increasing scale levels from fine to coarse.*

6. Conclusions and Future Work

In this work we presented a metric to guide visualization algorithms based on the wavelet transform and a model of the Human Visual System. The perceptual metric captures the change in images wrought by operators and the image synthesis algorithms. We would like to exploit the multiscale feature of the metric further and garner more insight into the images rendered at different scales. The perceptual weights were computed with the assumption that the wavelet transforms are ideal. Thus, it might be useful to recompute the weights. Also, redundant wavelet schemes like those of [16] cure some of the ills of the orthogonal wavelet transform, especially aliasing. It will be useful to learn about use of such a transform in developing a metric. Finally, the metric should be tested on more image synthesis algorithms to understand its performance completely.

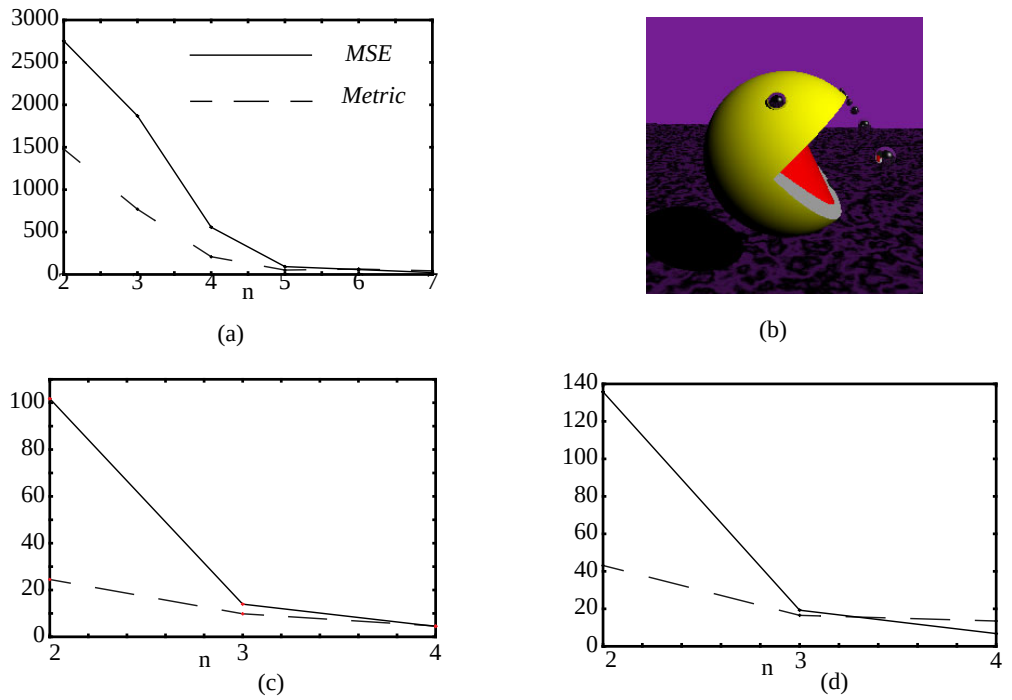

FIGURE 5. *Aliasing in Raytracing. In all incremental plots the MSE is the thick line, while perceptual metric is plotted with a dashed line. The averaging filter is a Gaussian for all experiments. The plots start with n=2 since no original exists. The y-axis includes both MSE and perceptual metrics. (a) Scheme with no jittering used on Pipes image. (b) PacMan Image (c) Supersampling scheme with no jittering on PacMan image (d) Supersampling with jittering on PacMan image.*

References

1. Chiu K., Herf M., Shirley P., Swamy S., Zimmerman K., "Spatially Nonuniform Scaling Functions for High Contrast Images", *Proceedings of Graphics Interface'93, May 1993*, pp. 245-254.
2. Cosman P. C., Gray R. M., Olshen R. A., "Evaluating Quality of compressed medical images: SNR, subjective rating and diagnostic accuracy", *Proceedings of the IEEE*, 82(6):919-932, June 1994.
3. Daubechies, I., "Ten Lectures on Wavelets", *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, PA, 1992.
4. Donoho, D. L., Johnstone, I. M., "Ideal Spatial Adaptation via Wavelet Shrinkage", *Biometrika* 81:422-455.
5. Furuie S., Herman G. T., Narayan T. K., Kinahan P. E., Karp J. S., Lewitt R. M., Matej S., "A methodology for testing statistically significant differences between fully 3D PET reconstruction algorithms", *Phys. Med. Biology*, 39:341-354, 1994.
6. Higgins G. C., "Image Quality Criteria", *Journal of Applied Photographic Engineering*, 1977, 3(2):53-60.
7. Jacobs, C. E., Finkelstein, A., Salesin, D. H., "Fast Multiresolution Image Querying", *Computer Graphics* (Proceedings of Siggraph'95), pp. 277-286.
8. Kim Y., Cho I., Lee I, Yun T., Park K. T., "Wavelet transform image compression using human visual

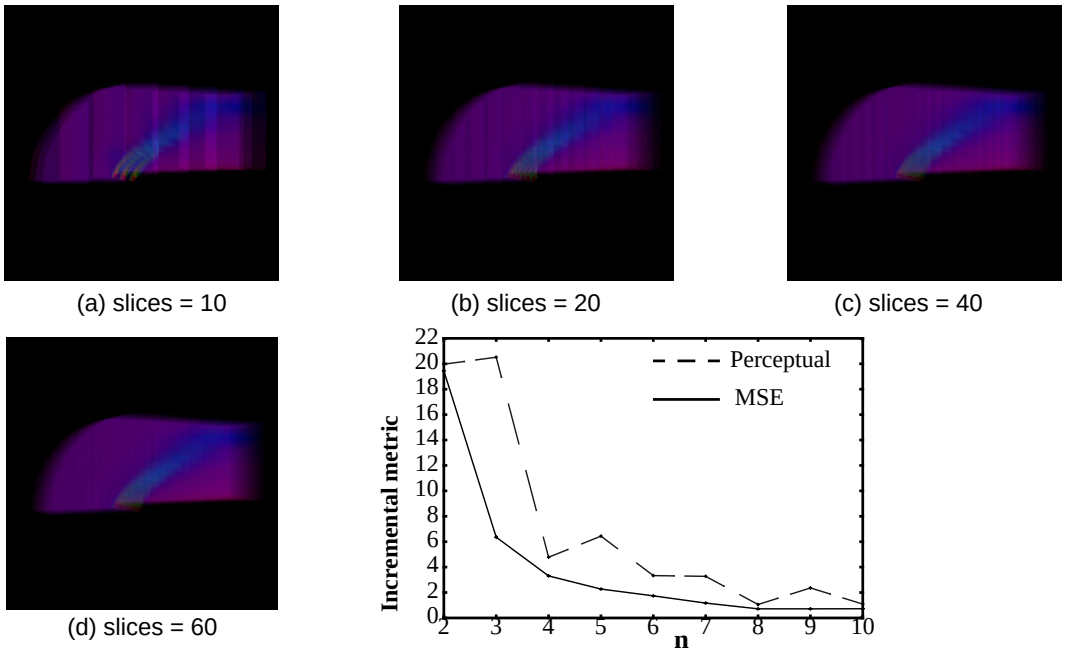

FIGURE 6. *Volume rendering algorithm for unstructured grids. Images for a varying number of slices are shown. The MSE (solid) and perceptual (dashed) metric plots indicate different behavior. Here, n >= 2 denotes difference between rendering using n x10 slices and (n-1)x10 slices.*

characteristics and a tree structure with a height attribute", *Optical Engineering*, Vol. 35, No.1, January 1996, pp. 204-212.

9. Levoy M., "Display of Surfaces from Volume Data", *IEEE Computer Graphics and Applications,* May 1988, 8(5):29-37.
10. Mannos, J. L., Sakrison, D. J., "The Effects of a Visual Fidelity Criterion on the Encoding of Images", *IEEE Transactions on Information Theory*, 20(3):525-536, July 1974.
11. Machiraju, R., Gaddipati, A., Yagel, R., "Steering Image Generation With Wavelet Based Perceptual Metric", presented at *IEEE Visualization '96* as a hot topic paper. (Technical Research Report, Department of Computer and Information Science, The Ohio State University, OSU-CISRC-6/96-TR)
12. Machiraju, R., Gaddipati A., Yagel, R, "Detection and Enhancement of Scale Coherent Structures Using Wavelet Transform Products", *Technical Conference on Wavelet Applications in Signal and Image Processing V,* SPIE Annual Meeting,San Diego, CA, 27 July - August 1, 1997 (to appear).
13. Mallat, S., "A Theory for Multiresoultion Signal Decomposition: The Wavelet Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674-693, July 1989.
14. Navarro P., Cristobal, G, "Image Analysis using Gabor Transform", *Advances in Electronics and Electron Physics*, (P.W. Hawkes ed.), Academic Press, San Diego, CA, Vol 8:(309-404).
15. Rushmeier H., Ward G., Piatko C., Sanders P., Rust B., "Comapring Real and Synthetic Images: Some Ideas About Metrics", Sixth Eurographics Workshop on Rendering, Dublin, Ireland, 1995, pp. 82-91.
16. Simoncelli E. P., Freeman W. T., Adelson E. H., Heeger D. J., "Shiftable Multi-Scale Transforms [or What's Wrong with Orthonormal Wavelets]", *IEEE Trans. Information Theory, Special Issue on Wavelets*, March 1992., 38(2):587-607.
17. Patrick T., Heeger D.,"Perceptual Image Distortion", *First IEEE International Conference on Image Processing*, vol 2, pp 982-986, November 1994.

18. Watson A. B., “The Cortex Transform: Rapid Computation of Simulated Neural Images”, *CVGIP*, 39:311-327, 1987.

19. Watson A. B., Gloria Y. Y., Joshua A. S., Villasenor J., “Visual thresholds for wavelet quantization error”, *Human Vision and Electronic Imaging*, Editors, Rogowitz B., Allebach J., Proceedings of the SPIE 2657, 382-392.

20. Watson A. B., Gale A. P., Joshua A. S., Ahumada A. J., “Visibility of DCT quantization noise: Effects of display resolution”, *Proceedings of Society for Information Display*, San Jose, CA, Society for Information Display, pp. 697-700

21. Wilson, T. A., Rogers, S. K., Myers, L. R., “ Perceptual Based Hyperspectral Image Fusion using Multiresoultion Analysis”, *Optical Engineering*, November 1995.

22. Witkin A., “Scale Space Filtering”, *Proceedings of International Joint Conference on Artificial Intelligence*, Karlsruhe, 1983.

23. Xu Y., Weaver J. B., Healy D. M., Lu J., “Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique”, *IEEE Transactions On Image Processing*, Vol. 3, No. 6, November 1994, pp. 747-758.

24. Yagel R., Reed D. M., Law A., Shih P., Shareef N., “Hardware Assisted Volume Rendering of Unstructured Grids by Incremental Slicing”, *Proceedings 1996 Symposium on Volume Visualization*, San Francisco, CA, September 1996, pp. 55-62.