

A Visual Difference Metric for Realistic Image Synthesis

Mark R. Bolin and Gary W. Meyer

Department of Computer and Information Science
University of Oregon
Eugene, OR 97403

ABSTRACT

An accurate and efficient model of human perception has been developed to control the placement of samples in a realistic image synthesis algorithm. Previous sampling techniques have sought to spread the error equally across the image plane. However, this approach neglects the fact that the renderings are intended to be displayed for a human observer. The human visual system has a varying sensitivity to error that is based upon the viewing context. This means that equivalent optical discrepancies can be very obvious in one situation and imperceptible in another. It is ultimately the perceptibility of this error that governs image quality and should be used as the basis of a sampling algorithm.

This paper focuses on a simplified version of the Lubin Visual Discrimination Metric (VDM) that was developed for insertion into an image synthesis algorithm. The simplified VDM makes use of a Haar wavelet basis for the cortical transform and a less severe spatial pooling operation. The model was extended for color including the effects of chromatic aberration. Comparisons are made between the execution time and visual difference map for the original Lubin and simplified visual difference metrics. Results from the realistic image synthesis algorithm are also presented.

Keywords: visual difference metric, image synthesis, adaptive sampling

1. INTRODUCTION

Realistic image synthesis involves the creation of a picture from a mathematical description of the world. Objects in the environment are modeled, a synthetic camera is placed in the scene, and the transport of light is simulated. Discrete samples are taken of the light energy at the picture plane, and the final image is reconstructed from these samples. The process is accomplished entirely within a computer. Real objects, light sources, and cameras are not required. The algorithms have become so sophisticated that the final result is often indistinguishable from a photograph.

These image synthesis techniques can be extended to take human perception into account. The formation of the image can be controlled by noting the places where improvements would be visually significant. Additional effort can then be invested to refine these parts of the picture. This is an adaptive sampling technique that utilizes a perceptual instead of an objective error metric. Image processing models of the visual system can be used to decide where additional samples of the environment should be taken.

In this paper a perceptually based image synthesis algorithm is described. An efficient implementation of a visual difference metric is used to direct the placement of samples as a picture is created. This work was first introduced in reference 3. Additional information is provided in Section 1 of this article regarding the visual difference metric that was utilized. The adaptive sampling algorithm is briefly described in Section 3. New results obtained by using the approach are presented in Section 4 and concluding remarks are made in Section 5.

2. VISUAL DIFFERENCE METRIC

Two of the most comprehensive image quality metrics are the Visual Difference Predictor (VDP) by Daly⁵ and the Sarnoff Visual Discrimination Metric (VDM) by Lubin.¹³ A recent study by Li^{11,12} compared the results of these two metrics. In this study it was found that although the Sarnoff VDM required somewhat more memory, it executed faster and produced better difference predictions. Another advantage of the Sarnoff model is its use of

Current address for Mark Bolin: Eastman Kodak Company, 1700 Dewey Avenue, Floor 1, Bldg. 65, Rochester, New York 14650-1816

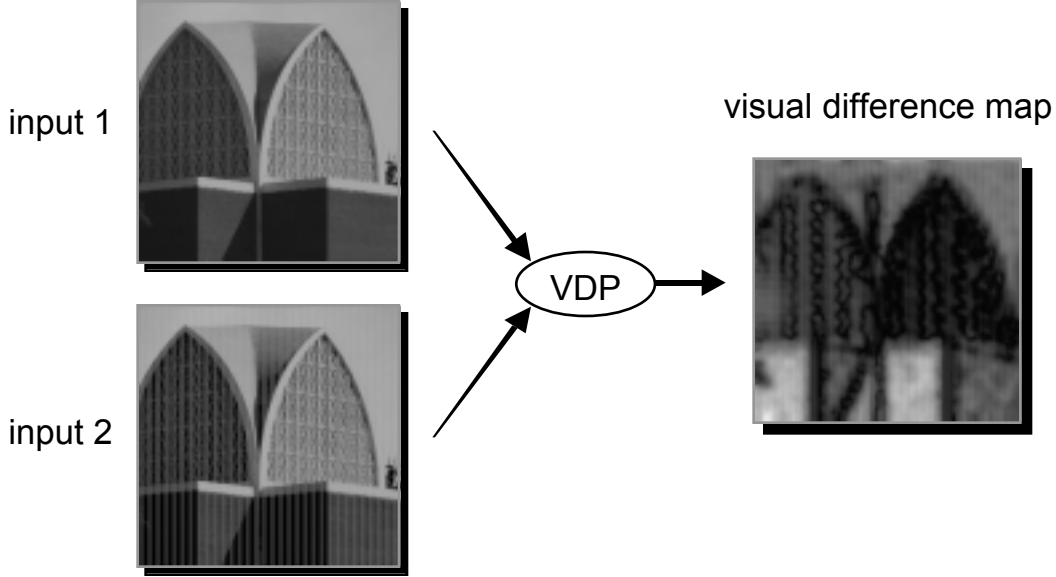


Figure 1. Input and output of the visual difference predictor.

a pyramidal transformation to isolate spatial frequency and orientation selective channels. The nature of this type of transform offers substantial efficiency benefits as will be seen in the adaptive sampling algorithm described in Section 3. For these reasons, the Sarnoff VDM was selected as a starting point for the development of the quality model discussed in this section. The new model has been modified to run efficiently, and it has been extended to handle color. This color extension is necessary because the original Sarnoff metric was only designed for achromatic images.

The input and output of the Sarnoff predictor are illustrated in Figure 1. In this example input 1 contains a chapel image, and input 2 is the same image distorted by an equal energy sinusoidal grating. It should be apparent that while the grating is uniform, its perceptibility is not. The distortion is most visible in the dark areas at the base of the chapel and less perceptible in the bright regions at the top of the image. The grating is also completely invisible inside the upper right archway because the lattice work in this area hides, or masks, the detectability of the grating. The output of the predictor is shown in the visual difference map on the right side of the figure. This image utilizes increased brightness to indicate areas with more perceptible differences as measured in terms of just noticeable differences (JND's). The difference map can be seen to have a good correspondence with a subjective comparison of the two inputs.

In this section the stages of processing involved in this visual difference predictor will be discussed. A block diagram of the model is given in Figure 2. This diagram illustrates the various processing steps that are involved. Each input image is independently passed through the steps labelled *cone fundamentals* through *spatial pooling*. The differences between the two images are accumulated in the *distance summation* step.

The input image is first encoded into the responses of the short (S), medium (M) and long (L) receptors found in the retina of the eye. This happens in the first stage of the vision model labelled *cone fundamentals*. The transformation used to convert from CIE XYZ space to SML space employs the following matrix equation¹⁴:

$$\begin{bmatrix} S \\ M \\ L \end{bmatrix} = \begin{bmatrix} 0.0000 & 0.0000 & 0.5609 \\ -0.4227 & 1.1723 & 0.0911 \\ 0.1150 & 0.9364 & -0.0203 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \quad (1)$$

The next step in the model is to apply a *cortex filtering* operation. The decomposition of an image into spatial frequency and orientation tuned channels is the most expensive operation performed by a visual model. Therefore, in order to significantly improve the execution time of a model, a high speed transform must be selected. The choice of this transform should also be influenced by the desire to incorporate the quality model within an adaptive

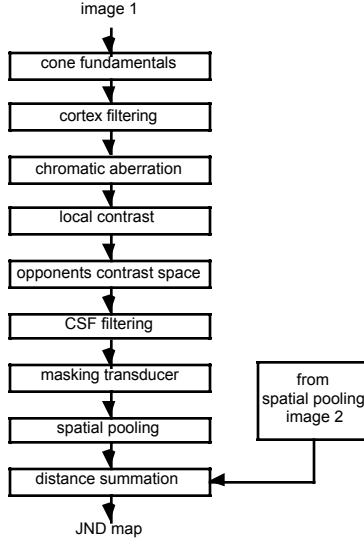


Figure 2. Block diagram of the visual difference predictor.

sampling algorithm. During the progression of an adaptive algorithm it is necessary to make numerous, iterative judgements about the quality of an image. Successive judgements are often made after modifying only a small region of the image. It would therefore be advantageous if small image modifications had a limited effect on the cortex representation, and if this effect could be rapidly calculated.

In order to satisfy these requirements, the Haar wavelet transform was selected to model the spatial frequency and orientation selectivity of the human visual system. This transform provides the fastest mechanism capable of decomposing an image into these selective channels. The Haar transform can be computed in $O(N)$ time, and, as will be shown in Section 3, it can be updated in $O(\log N)$ time during the progression of an adaptive sampling algorithm.

A number of other transforms were considered for this stage of the model. The cortex transform by Watson¹⁹ was one option. The disadvantage of this method is that it is based on a Fourier transform of the image. This transform requires $O(N \log N)$ time to compute. In addition, iterative refinement is slow because modifying the intensity of a single pixel affects all of the terms in this representation. A variety of other pyramidal transforms were also investigated. These included the steerable pyramid used in the Sarnoff model,¹³ Daubechies' family of wavelets,⁶ and the biorthogonal bases of Cohen, *et. al.*⁴ These methods were deemed undesirable because of the larger, overlapping spatial filters that are used in the transforms. The size of these filters slows down both the direct and iterative calculation of the transform, and the fact that the filters overlap would have complicated the error estimation stage of the adaptive sampling algorithm discussed in Section 3.

The Haar transform employed is the two-dimensional non-standard decomposition. This transform can be expressed as:

$$\begin{aligned}
 c_{l-1}\left[\frac{x}{2}, \frac{y}{2}\right] &= (c_l[x, y] + c_l[x + 1, y] + c_l[x, y + 1] + c_l[x + 1, y + 1])/4 \\
 d_{l-1}^1\left[\frac{x}{2}, \frac{y}{2}\right] &= (c_l[x, y] - c_l[x + 1, y] + c_l[x, y + 1] - c_l[x + 1, y + 1])/4 \\
 d_{l-1}^2\left[\frac{x}{2}, \frac{y}{2}\right] &= (c_l[x, y] + c_l[x + 1, y] - c_l[x, y + 1] - c_l[x + 1, y + 1])/4 \\
 d_{l-1}^3\left[\frac{x}{2}, \frac{y}{2}\right] &= (c_l[x, y] - c_l[x + 1, y] - c_l[x, y + 1] + c_l[x + 1, y + 1])/4,
 \end{aligned} \tag{2}$$

where c_l specifies the lowpass coefficients of the level l Haar basis, d_l^1 , d_l^2 and d_l^3 are the detail coefficients of the three two-dimensional level l Haar wavelets, and $c_{levels-1}[x, y]$ corresponds to the response of either the small, medium or long receptors at a pixel location (where *levels* represents the number of levels in the quad-tree used to store the Haar decomposition).

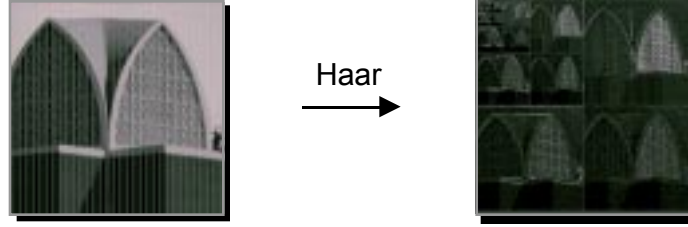


Figure 3. Flow graph of the *cortex filtering* stage of the visual difference predictor.

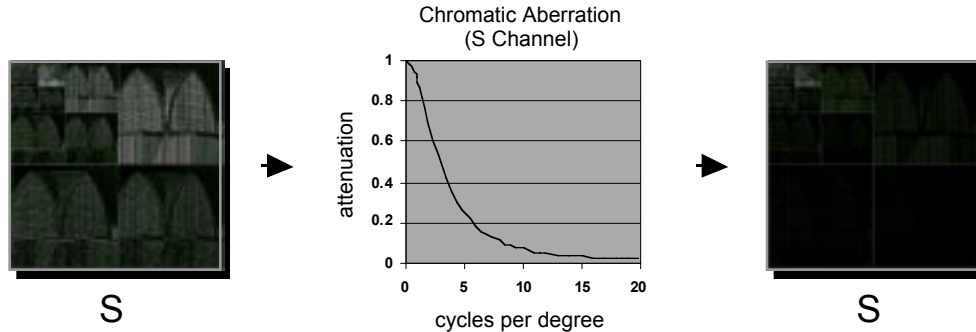


Figure 4. Flow graph of the *chromatic aberration* stage of the visual difference predictor.

Figure 3 illustrates the Haar transform applied to one of the SML color channels. In the image on the right, increased brightness indicates larger detail magnitudes. The highest frequency terms are arranged in the bottom and right side of the image and the lowest frequency term is in the upper left. At each level there are three blocks of detail coefficients. The top right, lower left, and lower right blocks respectively contain the horizontal (d^1), vertical (d^2), and diagonally (d^3) selective terms. The frequency selectivity of the detail terms at a given level of the representation is defined as the frequency in cycles per degree (cpd) to which the wavelet at that level is optimally responsive. The orientation and frequency selectivity of the Haar transform is a limitation of this approach. There are significant efficiency gains, however.

The next step in the image quality model incorporates the effect of *chromatic aberration*. This model is novel in its inclusion of this effect. Chromatic aberration describes the defocusing of light as a function of wavelength by the optics of the eye. This defocusing most strongly affects the response of the short wavelength receptors, and severely attenuates the visibility of high spatial frequency detail in this channel. The center graph of Figure 4 depicts a plot of the sensitivity loss due to chromatic aberration in the short wavelength channel. This plot shows that sensitivity drops to less than half its original value at 4 cpd and is virtually non-existent at frequencies higher than 8 cpd. The original chromatic contrast sensitivity experiments performed by Mullen¹⁵ corrected for chromatic aberration. In order to accurately apply the results of her work at the latter stages of the model it is necessary to reintroduce this effect.

Figure 4 illustrates how chromatic aberration is included in the image quality model. The unmodified cortex representation of the S cone receptors is illustrated on the left side of the figure. The response of these receptors are attenuated by the effect of chromatic aberration as a function of spatial frequency. The lowpass filter used is contained in the center graph. This filter was generated by a fit to the data of Marimont and Wandell.¹⁶ The lowpass filtering operation can be performed very rapidly because the cone responses are stored in a frequency based representation. Filtering in this domain is accomplished by merely scaling the detail coefficients by the amount of attenuation at the associated spatial frequency. The decreased response at high spatial frequencies can be seen in the resulting image on the right of the figure.

The eye's non-linear response to light is modeled in the stage labeled *local contrast*. The standard cone contrast calculation of $\frac{\Delta S}{S}$, $\frac{\Delta M}{M}$, and $\frac{\Delta L}{L}$ is accomplished by dividing the detail coefficients of each cone channel by the

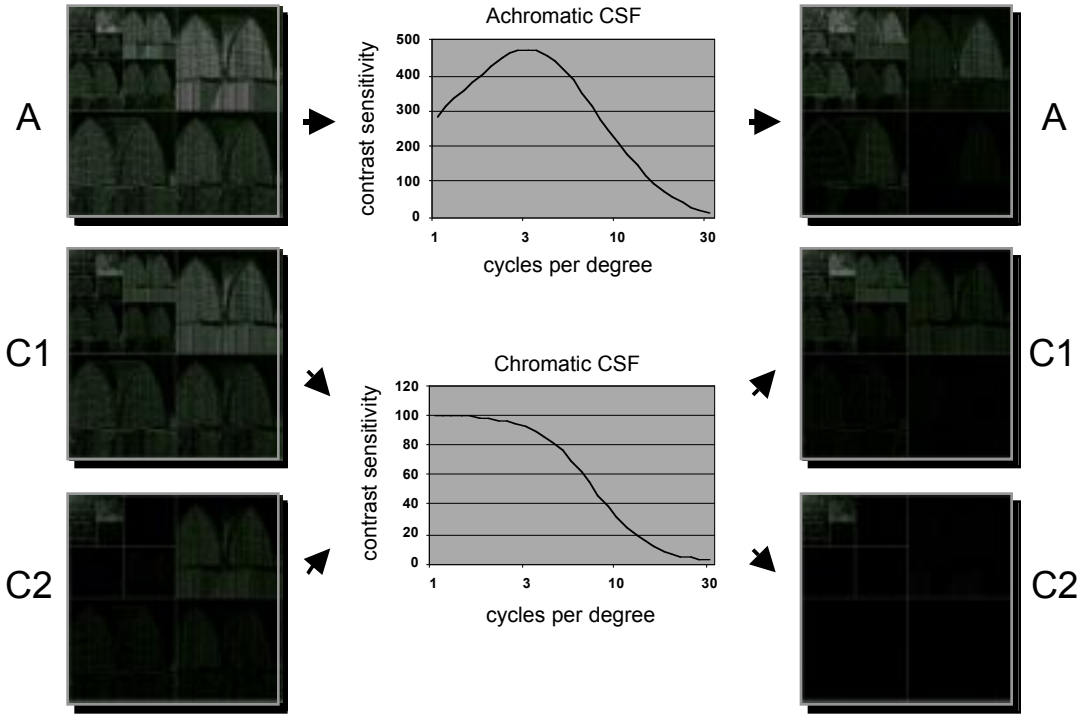


Figure 5. Flow graph of the *CSF filtering* stage of the visual difference predictor.

appropriate lowpass coefficient one level up in the quad-tree. This avoids the assumption, used in other models,^{5,7} that the eye can adapt at each pixel.

The *opponents contrast space* stage of the model comes next. The conversion of the cone contrasts to an opponents contrast space is accomplished using the transformation matrix¹⁴:

$$\begin{bmatrix} A \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 0.0001 & 0.2499 & 0.7647 \\ 0.0018 & 2.9468 & -2.5336 \\ 1.0111 & -0.3877 & 0.2670 \end{bmatrix} \begin{bmatrix} S \\ M \\ L \end{bmatrix}. \quad (3)$$

This equation shows that the achromatic channel is primarily determined by the combined responses of the medium and long wavelength receptors, the C_1 channel is composed of the difference in the responses of the medium and long wavelength receptors, and the C_2 channel largely isolates the responses of the short wavelength receptors.

A diagram of the sixth stage, labeled *CSF filtering*, is contained in Figure 5. Different contrast sensitivity functions are used for the achromatic and chromatic channels. For the achromatic channel the human visual system has a peak sensitivity to signals of around 4 cpd, and significantly less sensitivity at higher and lower spatial frequencies. The model uses the equation for the achromatic contrast sensitivity function that was presented by Barten.¹ For the chromatic channels visual sensitivity is strictly lowpass, with a lower peak sensitivity and a lower frequency cutoff than is present in the achromatic channel. The chromatic contrast sensitivity function that is used in the model is implemented with a Butterworth filter that has been fit to the chromatic sensitivity data from Mullen.¹⁵ The square of the contrast for each of the A , C_1 and C_2 channels is multiplied by the square of that channel's contrast sensitivity as a function of spatial frequency. The square of the contrast and contrast sensitivity function is used to model the energy response that occurs for complex cells, as described in the Sarnoff VDM.

The images on the right side of Figure 5 show the results of applying the achromatic and chromatic contrast sensitivity functions to the opponent contrast images. In the achromatic image, contrast response has been attenuated for both low and high spatial frequencies. For the chromatic channels contrast response declines with increasing spatial frequency. The fact that the C_2 channel has a lower frequency cutoff than the C_1 channel is the result of attenuation due to chromatic aberration that was modeled at an earlier stage of the algorithm.

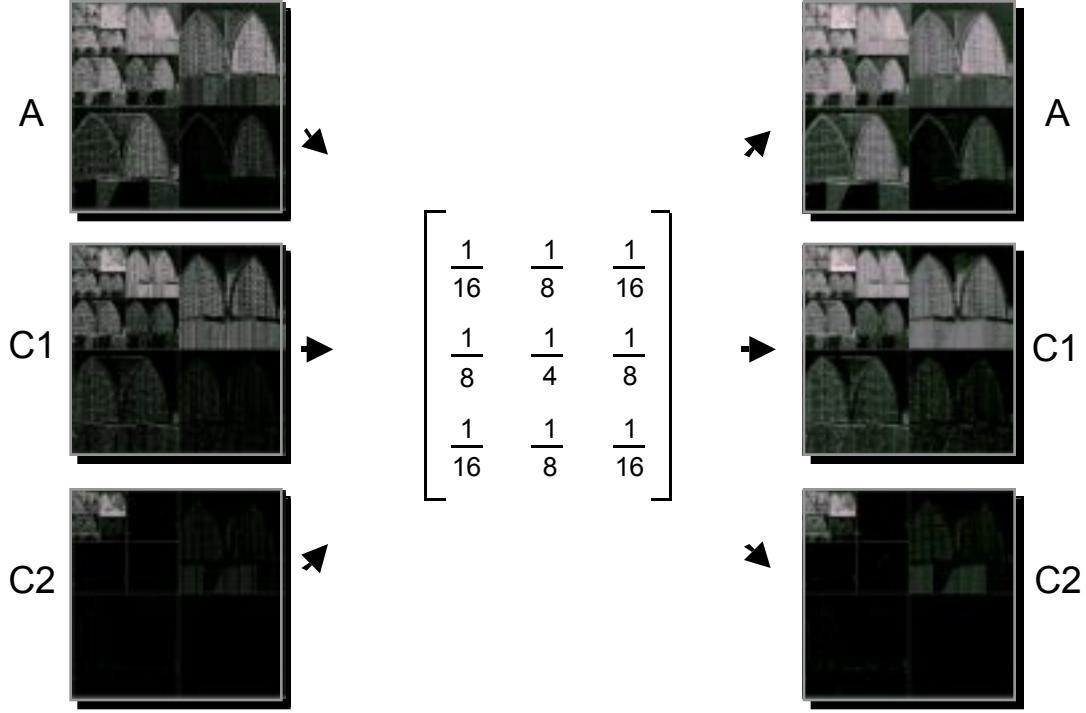


Figure 6. Flow graph of the *spatial pooling* stage of the visual difference predictor.

Visual masking is handled in the stage of the model labeled *masking transducer*. This property of the visual system, whereby strong signals of a given color, frequency, or orientation minimize the visibility of similar signals, is incorporated by using the same non-linear, sigmoid transducer that was employed in the Sarnoff VDM:

$$T(A) = \frac{2A^{2.25/2}}{A^{2.05/2} + 1}. \quad (4)$$

In this equation, $T(A)$ is the transducer output and A is the weighted contrast produced from the previous stage of the model. This transducer is applied independently to the contrasts of each of the A , C_1 , and C_2 color channels. This function augments low contrasts and compresses high contrasts. The net result is that differences between high contrast signals are reduced, whereas low contrast differences are increased. This simulates the masking and facilitation effect described by Legge and Foley.¹⁰

The inputs and outputs of the *spatial pooling* stage of the model are illustrated on the left and right sides of Figure 6 respectively. In this stage the transducer outputs are filtered to incorporate the fact that foveal human sensitivity is greater for sine wave gratings containing multiple cycles than it is for single cycle gratings. The pooling filter used in this model is contained in the center of Figure 6. The decision to use a 3x3 filter rather than the 5x5 filter specified in the Sarnoff VDM was made to improve the speed of the algorithm. This filter also corresponds better with the results of Wilson,²⁰ who indicated that sensitivity reaches its peak for gratings containing 2.5 cycles.

A visual difference map is computed in the final *distance summation* stage of the model. The local visual difference (LVD) at each node of the quad-tree is defined to be the sum across all orientations (θ) and color channels (c) of the differences of the pooling stages (P_1 and P_2) of the two images raised to the 2.4 power:

$$LVD = \sum_{\theta=1}^3 \sum_{c=1}^3 (P_1[\theta, c] - P_2[\theta, c])^{2.4}. \quad (5)$$

The final difference map is generated by accumulating local visual differences across levels. This is accomplished by summing the local difference down each path in the quad-tree and storing the result in the leaves. The output of the algorithm is given by the leaf differences raised to the 1/2.4 power. This distance summation stage is an application



Figure 7. The final visual difference map.

of Quick’s vector summation technique with a 2.4 exponent.¹⁸ The resulting visual difference map is contained in Figure 7.

Comparison of Figure 7 with the visual difference map in Figure 1 illustrates the differences between the results obtained with the simplified model and the original Sarnoff VDM. In the simplified model, blocking artifacts are produced by the Haar wavelet decomposition. Aside from this difference, the results of both algorithms are similar, correspond well with a subjective comparison of the input images, and, as will be shown in Section 4, are usable in a realistic image synthesis algorithm. The simplified model also executed in $1/60^{th}$ of the time of the original Sarnoff metric. This is true even though the Sarnoff VDM processed one channel in a gray-scale image representation and the new model processed three color channels.

3. ADAPTIVE SAMPLING ALGORITHM

The vision model described in the preceding section has been integrated into a realistic image synthesis algorithm. This algorithm constructs an image in such a way that the visual difference metric can be evaluated while the picture is being created. This is accomplished by keeping track of the variance across the image and by constructing two boundary images from this statistical information. The visual difference metric is run on the boundary images and the results are used to determine where to take the next image sample. This section of the paper will briefly describe each of the steps in the adaptive sampling algorithm shown in Figure 8. Additional details can be found in reference 3.

The algorithm computes a Haar representation of the image in SML space as it takes samples of the environment. This is accomplished in the parts of the block diagram labelled *cone fundamentals* and *refine cortex representation*. A technique similar to the “splat and pull” method described in Gortler, *et. al.*⁸ is used to create the cortical representation. Each sample taken is averaged into the leaves of the quad tree that it affects. This includes the pixel at the most detailed level of the representation all the way up the quad tree to the lowest frequency terms that

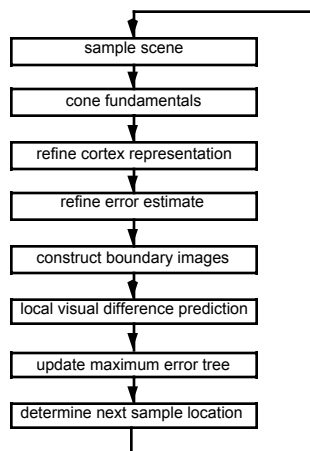


Figure 8. Block diagram of the basic adaptive sampling algorithm.

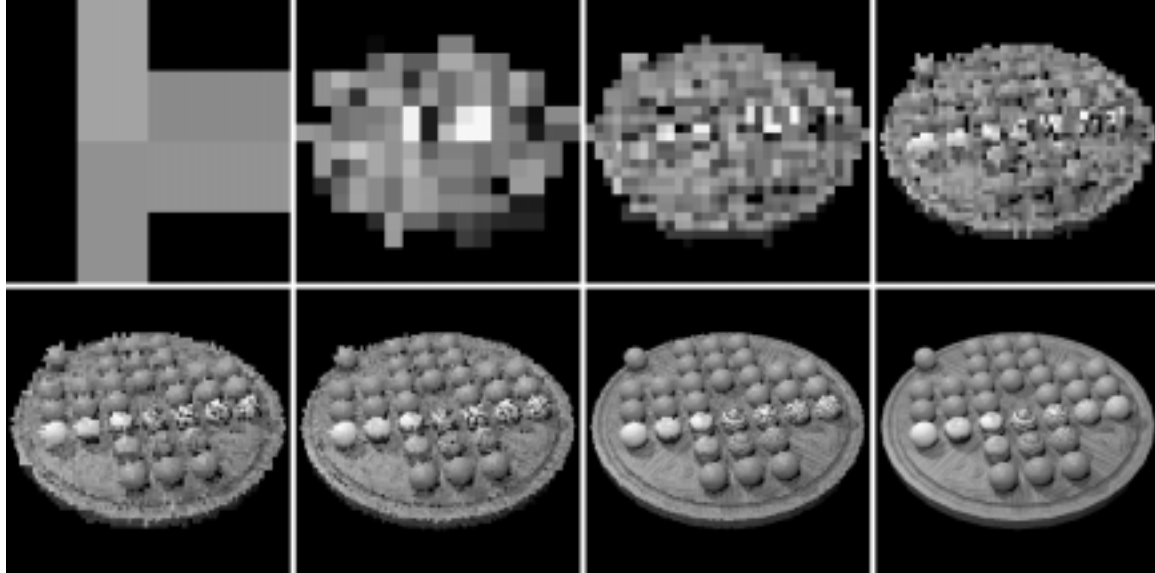


Figure 9. Illustration of the *refine cortex representation* stage. The number of samples used to construct these images increases from left to right, and from top to bottom.

contains this pixel. The fact that each sample only affects a few terms at each level of the quad tree is what makes this algorithm feasible. Figure 9 shows how the image looks as the cortex representation is formed.

In the next two steps of the algorithm, *refine error estimate* and *construct boundary images*, two images are created for evaluation by the visual difference metric. First the variance is determined for each of the lowpass and detail coefficients of the Haar representation. The variance at a leaf node is just the variance of each sample in that leaf divided by the number of samples.^{2,9} For an interior node in the quad tree the variance is equal to the sum of the variance of the four children divided by 16. Two boundary images are constructed from the current Haar representation and the variance information. The magnitude of the details for these two images are taken from the 25% and 75% points on a standard deviation curve with the spread of the curve determined by the variance.

The locations of perceptible error in the image are determined in the steps labeled *local visual difference prediction* and *update maximum error tree*. The two boundary images are passed through the stages of the visual difference metric labeled *local contrast* through *spatial pooling* in Figure 2. The local error stored at each node of the quad tree is as specified in the *distance summation* stage of the vision metric. The maximum error for a node keeps track of the largest visual difference below that point in the error tree. It is defined as the local error plus the largest maximum error in one of the four children. The largest visual error contained at any location in the image plane is the maximum error value stored in the root node of the quad tree. The maximum *total* visual difference at any node in the error tree can be found by performing a depth-first traversal of the tree. This value is easily calculated by summing the local visual difference contributions down the traversed path and adding this sum to the maximum visual difference stored at the node.

The placement of the next sample is done in the stage labeled *determine next sample location*. To find this position the quad tree can be traversed and the branch with maximum visual error taken until the bottom of the tree is reached. A better approach is to take a sample in all regions of the image that are above the user specified tolerance. These areas can be found by traversing the tree to locate all nodes at which the maximum total error is greater than the desired threshold. A sample is then taken in a part of the image that lies below each of these nodes in the quad tree. In addition, instead of acquiring a single sample in each of these areas, multiple samples can be taken. The number of samples to take is based upon the observation that the perceptual error declines with approximately the square-root of the sampling rate (see Figures 12 and 13). Sampling all regions that need work and taking multiple samples in these areas minimizes the number of times that the expensive image quality metric must be evaluated. The stopping condition for the algorithm occurs when the maximum error in the root node drops below a specified tolerance.

4. RESULTS

This section compares the results of the new perceptually based adaptive sampling algorithm with the results of two commonly used sampling strategies. This comparison will cover both the quantity of samples that are required to produce images of a given visual quality and the overall expense of the algorithms. A number of example renderings will be used to demonstrate the key features of the perceptually based technique. The two other sampling strategies that will be employed in this comparison are uniform sampling and adaptive sampling based on an objective error estimate.

Uniform sampling is the simplest and therefore one of the most prevalent methods for placing samples within the image plane. In this technique a refinement test is not used and an equal number of samples are taken in each pixel. The method begins by taking a single sample at each screen location. This sampling is performed left to right, top to bottom across the image. After all pixels have been sampled once, a second sample is taken in each pixel. This process continues until the final sampling density has been reached. Within a given pixel the samples are randomly distributed and the intensity of a pixel is defined to be the average of the samples taken within it. One of the drawbacks of this strategy is that it is the responsibility of the user to determine the sampling rate that produces an image of the desired quality.

The second sampling strategy that will be used for comparison is adaptive sampling based on an objective error estimate. This algorithm uses the variance of the sample’s radiance in RGB color space as its error metric. This approach is similar to a number of the prior techniques that use sample statistics as the basis for their refinement test.^{9,17} The actual adaptive algorithm was created by removing the stages that modeled the human visual system from the basic perceptually based adaptive sampling strategy described in Section 3.

The objective adaptive sampling algorithm receives the sample’s RGB radiance as input. The goal of this algorithm is to iteratively place each sample at the location containing the largest objective error. This is accomplished by creating and refining a Haar wavelet image approximation and multi-resolution error estimate as described in the *refine cortex representation* and *refine error estimate* stages of the basic perceptual algorithm (see Section 3). The local error (*LOD*) at each node of the quad-tree is defined to be the sum of the detail variance (*V*) across all detail orientations (θ) and RGB color channels (*c*):

$$LOD = \sum_{\theta=1}^3 \sum_{c=1}^3 V[\theta, c]. \quad (6)$$

A maximum error tree is created by summing the largest local error up the branches of the quad-tree as described previously. The result of this operation is that a value is stored at each node of the tree that represents the largest variance present in the region of the image which that node is defined to cover. The next sample location is determined by traversing the quad-tree in a top-down fashion and selecting the node with the largest variance. In this manner samples are always placed in the location of the image plane containing the largest objective error.

4.1. Sampling Rates

This section will discuss the number of samples required by the different sampling strategies in order to produce images of a given visual quality. This discussion compares the results of the perceptual algorithm with the results of the uniform and objective adaptive sampling techniques described above. The examples show that the new perceptually based adaptive sampling algorithm is able to produce images of equivalent visual quality using fewer samples than either of the two existing sampling algorithms.

Two example renderings will be used to highlight situations where adaptive sampling with an objective metric leads to erroneous results. In each of the examples the placement of samples by the three approaches will be discussed. The images that are produced by the algorithms after an equivalent number of samples will be shown in order to allow a visual inspection. Additionally, a visual difference prediction will be performed using the algorithm described in Section 1 to compare the images with a high quality rendering. This will illustrate the areas of the image containing visible artifacts and further verify the new difference predictor. Finally, a graph of the maximum visual difference versus number of samples will be given for each of the three techniques. The maximum visual difference is defined to be the largest difference found at any location in the image, using the new visual difference predictor. A rendering is generally not considered to be of high enough quality until all regions of the image are computed accurately enough

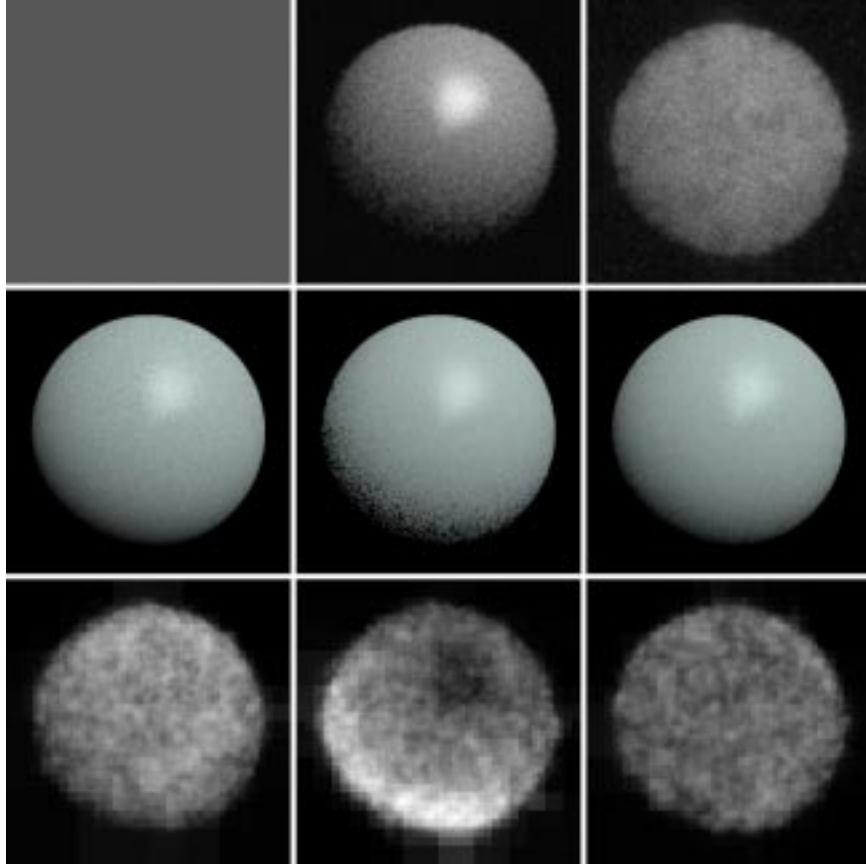


Figure 10. Comparison of uniform (left), objective (middle), and perceptual (right) sampling strategies for the contrast example. The rows contain the sample density maps (top), the images produced after an average of 20 samples per pixel (center), and the visual difference comparison with the high quality image (bottom).

so that the error is below the visual threshold. Therefore, if even a small region of the image contains significant perceptible error, the image can be considered unusable. For this reason maximum visual difference is an appropriate quantity in these comparisons.

In the first example we will see a situation where placing more samples in regions of large objective error will create images with more perceptible artifacts than if the same number of samples had been placed uniformly. This contradicts the common wisdom in computer graphics that adaptive algorithms based on objective error will always outperform uniform sampling. However, a perceptually based sampling algorithm can efficiently handle this situation.

The scene for this example consists of a simple sphere that is illuminated by a small area light source. Blind Monte Carlo integration is performed to evaluate the shading integral. This shading technique spawns many rays at random directions from each surface intersection in order to evaluate the radiance that is incident at a point on the sphere. Since this is a random process there is a certain amount of variance in the intensity of the samples taken of the sphere. Because there is very little spatial variation in this image, intensity variation and the visibility of this variation at different illumination levels are the primary factors that govern the appropriate sampling rate.

The sample density maps for the three sampling algorithms are shown in the top row of Figure 10. The map in the center shows the sample density for the objective method. In this image we see that the most samples are taken at the brightest regions of the sphere and the least samples in the dark regions. This is because the standard deviation of the samples scales with the reflectivity of the sphere. Consider for example that the scene is illuminated by a 100 cd/m^2 light source. At each point on the surface rays are spawned to determine the incident light. Rays that strike the light source will return the intensity of the light. Rays that miss the light source will return an intensity of 0 cd/m^2 . Therefore, samples that originate at the image plane and strike a point on the sphere that reflects 100%

of the incident light will return noisy values between 0 and 100 cd/m^2 , depending on the number of spawned rays that strike the light source. However, samples from the image plane that intersect a point on the sphere that only reflects $1/100^{\text{th}}$ of the incident light will only vary between 0 and 1 cd/m^2 (the difference in the reflectivity of these two points is the result of the orientation of the surface relative to the light source and eye position). Thus, the amount of noise in the first case will be 100 times greater than the amount of noise in the second case.

As it turns out, the sampling pattern produced by the objective algorithm is extremely inefficient. This is because the sensitivity of the human visual system varies with the local illumination level. The visual system is much more tolerant of error in bright regions than in dark and is equally tolerant of error when $\frac{\Delta L}{L}$ is a constant (where ΔL represents the luminance error and L is the mean luminance). In this example the mean luminance at locations of the image also scales with the reflectivity of the sphere. The net result is that the visibility of the error at a given sampling rate is uniform across the face of the sphere. This implies that uniformly sampling the interior of the sphere is an optimal solution. This is the sampling pattern used by the perceptual algorithm, as can be seen in the rightmost sample density map.

The middle row of Figure 10 shows the images created by the three algorithms after an average of 20 samples per pixel. Note that noise is very visible along the dark underside of the sphere in the image produced by the objective method, whereas it is difficult to discriminate the noise anywhere on the surface of the sphere produced by the perceptual method. Additionally, the image rendered by the perceptual algorithm has a somewhat higher visual quality than the image generated by the uniform sampling strategy. This is because the perceptual algorithm cast fewer samples in the constant background around the edges of the image and, instead, concentrated these samples in the interior of the sphere where they were most needed. These observations are further demonstrated by the visual difference maps contained in the bottom row of this figure.

The visual quality of the images produced by these algorithms is plotted versus the sampling rate in Figure 12. The objective sampling algorithm has the worst performance in this example. Because of the poor sample distribution used by this method a large number of samples are wasted in the bright specular region of the sphere before an adequate number of samples are taken in the darker regions. The uniform sampling algorithm produced better results, requiring only a quarter of the samples of the objective method. This is due to the fact that uniform sampling is exactly the right thing to do within the interior of the sphere. The perceptually based algorithm fared the best, requiring only half as many samples as the uniform method.

The second example demonstrates the effect of masking. This scene consists of two rectangles. The left rectangle reflects a uniform gray. The right rectangle is texture mapped with the top view of a section of carpet. This scene is illuminated with an area light source and blind Monte Carlo integration is performed to evaluate the shading integral. This process results in significant variation in the intensity of the samples at any given location.

The sample density maps for this example are depicted in the top row of Figure 11. The map produced by the objective method shows that more samples have been taken in the right rectangle than were taken in the left. This is because the additional spatial variation of the carpet creates a greater sample variance than in the left rectangle where there is only intensity variation caused by the Monte Carlo integration. However, this sampling pattern is inefficient. The texture map of the carpet contains significant energy at spatial frequencies to which the visual system has a high sensitivity. This energy is additionally distributed across a number of frequencies and orientations. The result of this energy distribution is that the white noise produced by the intensity variation is masked by the presence of the carpet texture. Therefore, an equivalent amount of noise will be less apparent on the texture mapped rectangle than on the uniform one, where no masking occurs. This effect is correctly incorporated by the perceptual algorithm which takes more samples in the left rectangle than in the right.

The images produced by the three algorithms after 10 samples per pixel are shown in the middle row. At this stage noise is still apparent in all of the images. Within the image produced by the objective algorithm the artifacts are the strongest and occur within the left rectangle where there is no masking. The right rectangle contains little perceptible error. The image produced by the uniform sampling distribution is somewhat better because an equal number of samples are taken in each rectangle. However, the error is still more apparent in the left rectangle than in the right. The image from the perceptually based algorithm is the only one with a visibly uniform error distribution. This approach has significantly more objective error in the textured rectangle than in the non-textured one. Due to the effect of masking, however, the two are of equivalent perceptual quality.

Visual difference maps of images produced by the algorithms are contained in the bottom row. The difference maps for the uniform and objective algorithms show non-uniformity in how visible the error is in the two rectangles.

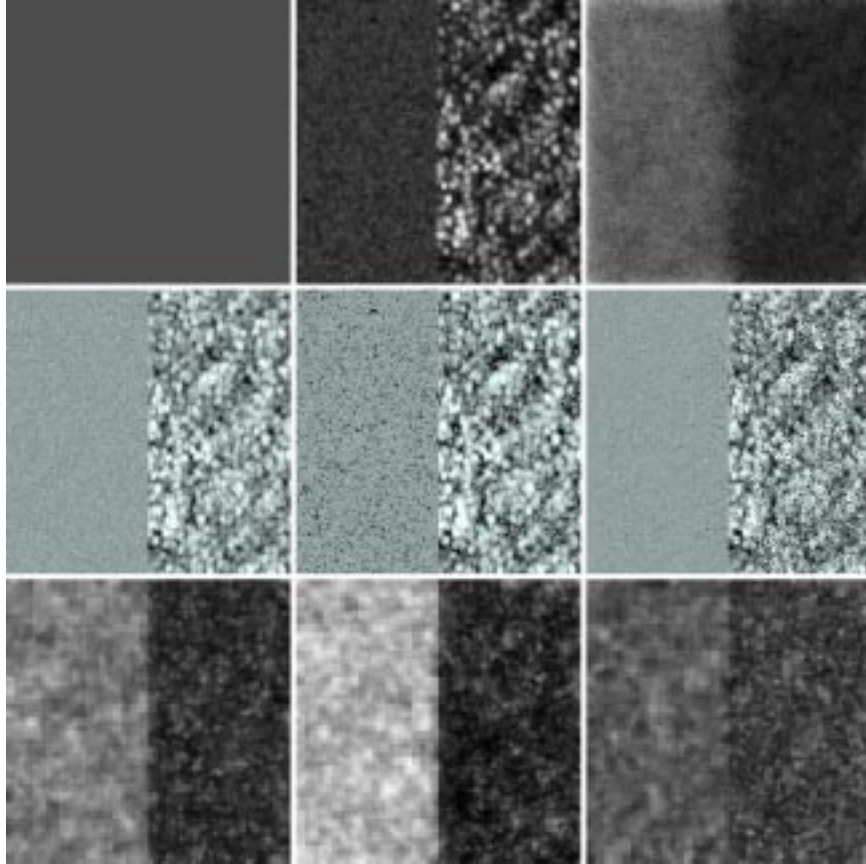


Figure 11. Comparison of uniform (left), objective (middle), and perceptual (right) sampling strategies for the masking example. The rows contain the sample density maps (top), the images produced after an average of 10 samples per pixel (center), and the visual difference comparison with the high quality image (bottom).

The difference map for the perceptual algorithm is more uniform. In the early stages of sampling, the difference maps for the perceptually based algorithm still exhibit some non-uniformity, with more error on the left rectangle than on the right. This occurs because a number of samples are required before the algorithm can ascertain a reasonable estimate of the spatial frequency spectrum.

Uniformity of perceptible error is a key idea in improving the performance of sampling algorithms. If the perceptibility of error is uniform across the image plane then the peak visual error has been minimized. This concept allows the perceptually based adaptive sampling algorithm to minimize the number of samples required to compute images to a given visual tolerance.

For the masking example, visual difference is plotted against number of samples in Figure 13. In this graph it is shown that the objective sampling algorithm required the most samples to accurately render the image, requiring roughly 1,000 samples per pixel. This extraordinarily high sampling rate was required because of the large sample variance and the inefficiency of the sampling pattern. Uniform sampling performed better in this instance, requiring a little over half as many samples. By correctly incorporating the effect of visual masking, the perceptual algorithm performed best of all, requiring only a third as many samples as the objective sampling algorithm.

4.2. Execution Time

This section presents the results of a number of timing tests that compare the perceptually based adaptive sampling technique with two previous sampling strategies. The perceptual algorithm is shown to produce images of a given visual quality in less time than is required by previous sampling methods.

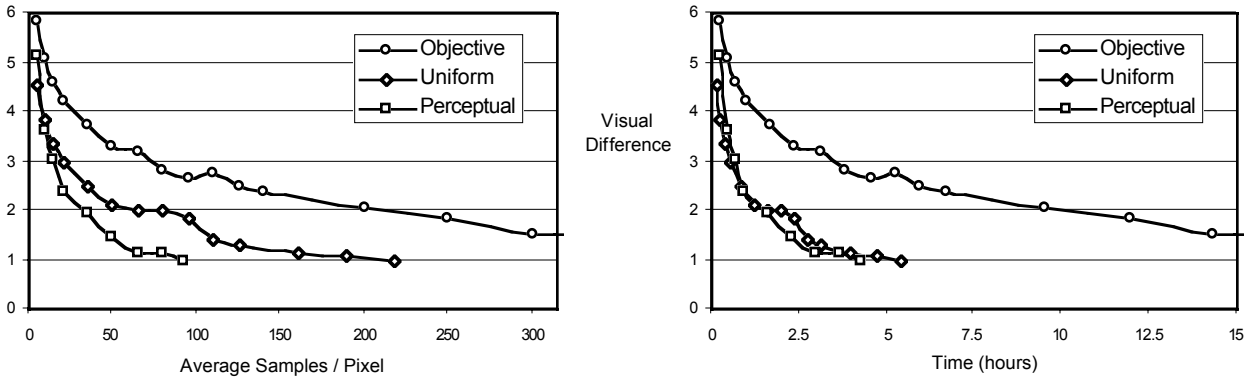


Figure 12. Sampling rates (left) and timing tests (right) for the contrast example.

The timing tests employ a similar technique to the one used to determine required sampling rates in the previous subsection. Two example scenes will be rendered using three different sampling algorithms. The images that are produced by these algorithms are output at specified intervals. The perceptual quality of the images is then computed by a comparison to a high quality rendering of the same scene. This comparison utilizes the visual difference predictor discussed in Section 1. These results are accumulated in a graph that plots the maximum perceptual difference between the two images versus length of execution time. This is essentially a remapping of the required sampling rate graphs along a time axis.

Figures 12 and 13 show the plots of maximum visual difference versus execution time for the scenes depicted in Figures 10 and 11 respectively. Observe the aggregate speed of the perceptual algorithm in comparison to the previous objective sampling technique. This comparison demonstrates the ultimate benefit of the new perceptually based adaptive sampling method. The previous subsection discussed the number of samples required by the objective method to produce an image of a given visual quality. This approach was shown to require three to ten times the number of samples necessary with the perceptually based technique. This savings in the sampling rate translates directly into a savings in overall execution time. In both examples, the perceptual algorithm is able to render images to the visible threshold using less time than both the uniform and objective sampling techniques.

Figure 14 accumulates the results of the above timing tests and two more tests into a single chart. The additional spatial frequency and chromatic spatial frequency tests are cases where uniform sampling does especially poorly. It is well known that an adaptive approach is superior in situations involving high frequencies and direct light source sampling. The spatial frequency and chromatic spatial frequency results for the objective and perceptual methods confirm this fact. The values in the chart are derived by rendering the images to the visible threshold (visual difference = 1 in the timing graphs), and the times are reported as a percentage of the worst case performer. In this chart we see that the perceptual algorithm is able to render the example images using only 10.0 to 28.1 percent of the elapsed

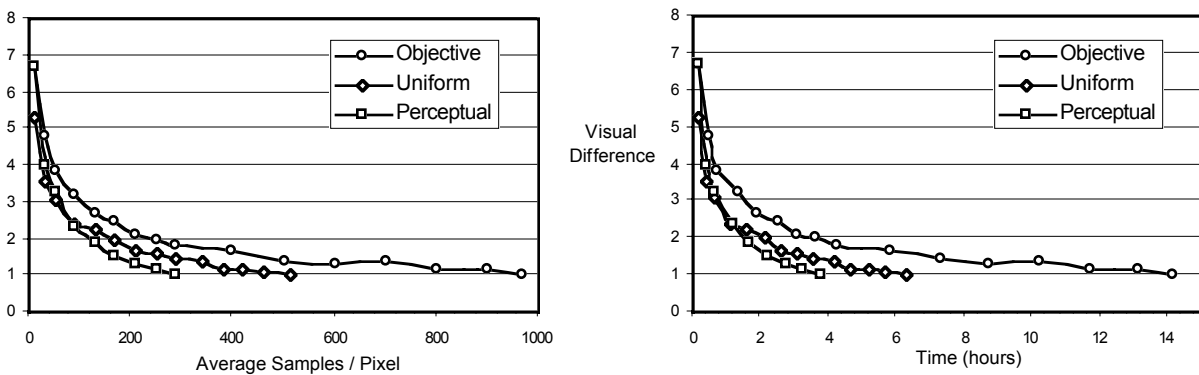


Figure 13. Sampling rates (left) and timing tests (right) for the masking example.

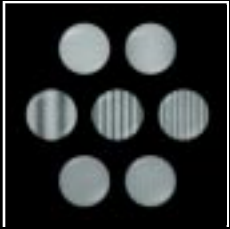
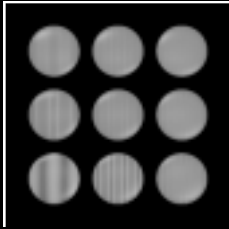

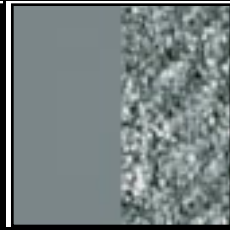
Relative Time				
Direct			Monte Carlo	
				
	Spatial Frequency	Chromatic Spatial Frequency	Contrast	Masking
Uniform	100.0	100.0	12.8	44.7
Objective	27.0	52.1	100.0	100.0
Perceptual	17.0	28.1	10.0	26.5

Figure 14. A summary of the timing test results. The values are based on the elapsed time required to render the images to the perceptual threshold. The time are reported as a percentage of the slowest algorithm.

time required by the existing algorithms. This is a significant decrease in execution time.

5. CONCLUSIONS

An efficient implementation of the Sarnoff VDM has been created. This version of the Sarnoff metric was extended to handle color and it includes the effects of chromatic aberration. The visual difference map produced by this algorithm compares favorably with the map generated by the Sarnoff VDM although some blocking artifacts are introduced due to the use of the Haar transform. Even though it has three color channels to process, the new method executes in $1/60^{th}$ of the time of the Sarnoff VDM.

This efficient visual difference metric was integrated into a realistic image synthesis algorithm. This made it possible to use a perceptual instead of an objective error metric to control an adaptive sampling algorithm. The perceptual algorithm was shown to preserve the superior behavior of adaptive sampling over uniform sampling in simple illumination situations containing high frequencies. The perceptual algorithm was also shown to improve the performance of adaptive sampling when high frequency color detail is present in these simple lighting cases. In more complex Monte Carlo lighting simulations, the perceptual algorithm produced better results than the adaptive approach with an objective error metric. This was shown to be true in cases involving both strong contrast and spatial masking effects. The tests show that while uniform and simple adaptive sampling fail in certain cases, the perceptual metric performs well in all situations.

ACKNOWLEDGMENTS

The authors would like to thank Jae H. Kim for his help in creating Figure 9. This research was funded by the National Science Foundation under grant number CCR 96-19967.

REFERENCES

1. Barten, P. G. J., "The Square Root Integral (SQRI): A New Metric to Describe the Effect of Various Display Parameters on Perceived Image Quality," *Human Vision, Visual Processing, and Digital Display*, Proc. SPIE, Vol. 1077, pp. 73-82, 1989.
2. Bolin, M. R. and Meyer G. W., "An Error Metric for Monte Carlo Ray Tracing," *Rendering Techniques '97*, J. Dorsey and P. Slusallek, Editors, Springer-Verlag, New York, pp. 57-68, 1997.
3. Bolin, M. and Meyer, G., "A Perceptually Based Adaptive Sampling Algorithm," *Computer Graphics, Annual Conference Series*, ACM SIGGRAPH, pp. 299-309, 1998.
4. Cohen, A., Daubechies, I., and Feauveau, J. C., "Biorthogonal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, Vol. 45, No. 5, pp. 485-500, 1992.
5. Daly, S., "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity," *Digital Images and Human Vision*, A. B. Watson, Editor, MIT Press, Cambridge, MA, pp. 179-206, 1993.
6. Daubechies, I., "Orthonormal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, Vol. 41, No. 7, pp. 909-996, 1988.
7. Ferwerda, J. A., Shirley, P., Pattanaik, S. N., and Greenberg, D. P., "A Model of Visual Masking for Computer Graphics," *Computer Graphics, Annual Conference Series*, ACM SIGGRAPH, pp. 143-152, 1997.
8. Gortler, S. J., Grzeszczuk, R., Szeliski, R., and Cohen, M. F., "The Lumigraph," *Computer Graphics, Annual Conference Series*, ACM SIGGRAPH, pp. 43-54, 1996.
9. Lee, M. E., Redner, R. A. and Uselton, S. P., "Statistically Optimized Sampling for Distributed Ray Tracing," *Computer Graphics, Annual Conference Series*, ACM SIGGRAPH, pp. 61-67, 1985.
10. Legge, G. E. and Foley, J. M., "Contrast Masking in human vision," *Journal of the Optical Society of America*, Vol. 70, pp. 1458-1470, 1980.
11. Li, B., "An Analysis and Comparison of Two Visual Discrimination Models," *Master's Thesis, University of Oregon*, June 1997.
12. Li, B., Meyer, G. W., and Klassen, R. V., "A Comparison of Two Image Quality Models," to appear in *Human Vision and Electronic Imaging III*, B. E. Rogowitz and T. N. Pappas, Editors, Proc. SPIE, Vol. 3299, 1998.
13. Lubin, J., "A Visual Discrimination Model for Imaging System Design and Evaluation," *Vision Models for Target Detection and Recognition*, Eli Peli, Editor, World Scientific, New Jersey, pp. 245-283, 1995.
14. Meyer, G. W., "Wavelength Selection for Synthetic Image Generation," *Computer Vision, Graphics, and Image Processing*, Vol. 41, pp. 57-79, 1988.
15. Mullen, K. T., "The Contrast Sensitivity of Human Colour Vision to Red-Green and Blue-Yellow Chromatic Gratings," *J. Physiol. (Lond.)*, Vol. 359, pp. 381-400, 1985.
16. Marimont, D. H. and Wandell, B. A., "Matching Color Images: The Impact of Axial Chromatic Aberration," *J. Opt. Soc. Am. A*, Vol. 12, pp. 3113-3122, 1993.
17. Painter, J. and Sloan, K., "Antialiased Ray Tracing by Adaptive Progressive Refinement," *Computer Graphics, Annual Conference Series*, ACM SIGGRAPH, pp. 281-288, 1989.
18. Quick, R. F., "A Vector-Magnitude model for Contrast Detection," *Kybernetik*, Vol. 16, pp. 65-67, 1974.
19. Watson, A. B., "The Cortex Transform: Rapid Computation of Simulated Neural Images," *Computer Vision, Graphics, and Image Processing*, Vol. 39, pp. 311-327, 1987.
20. Wilson, H. R., "Psychophysical Models of Spatial Vision and Hyperacuity," *Vision and Visual Dysfunction: Vol. 10: Spatial Vision*, D. Regan, Editor, CRC Press Inc., Boston, MA, pp. 64-86, 1991.