The Visible Differences Predictor: applications to global illumination problems

Karol Myszkowski
The University of Aizu; Aizu-Wakamatsu 965-8580; Japan

Abstract

In this study of global illumination computations, we investigate the applications of the perceptually-based Visual Difference Predictor (VDP) developed by Daly [5]. First, we validate the performance of this predictor in shadow masking by texture and luminance contrast experiments. We also experiment with Contrast Sensitivity Functions (CSFs) derived from the results of various psychophysical experiments, various spatial frequency and orientation channel decomposition schemes, and contrast definitions, in order to check predictor integrity and sensitivity to differing models of visual mechanisms. We show applications of the VDP to monitor the perceived quality of the progressive radiosity and Monte Carlo solutions, and decide upon their stopping conditions. Also, based on the local error metric provided by the predictor we show some initial attempts to drive adaptive mesh subdivision in radiosity computations.

1 Introduction

Since the beginning of computer graphic studies, the quest for development of perception-driven techniques has been and continues to be an important issue. This is mostly because of the expected dramatic performance gains by focusing computations on scene features which can be readily perceived by human observers under given viewing conditions. This seems especially attractive for global illumination solutions [10], which deal with very costly computations of light interactions between the surfaces modeling a scene. Current algorithms usually rely on energy-based metrics of solution errors, which do not necessarily correspond to the visible improvements of the image quality [13]. Ideally, one may advocate the development of perceptually-based error metrics which can control the accuracy of every light interaction between surfaces. This can be done by predicting the visual impact those errors may have on the perceived fidelity of the rendered images. In practice, apart from the conceptual difficulties in elaborating such a low level error metric, it could be costly in evaluation, and could depend strongly upon a specific global illumination algorithm.

Another approach is to develop a perceptual metric which operates directly on the rendered images. Of course, such a metric provides a summary of the algorithm performance as a whole rather than giving a detailed insight into the work of its particular elements. However, the metric and *a priori* knowledge of the current stage of computations can be used to obtain more specific measures of adaptive meshing performance, accuracy of shadow reconstruction, convergence of the solution for indirect lighting, and so on. If the goal of rendering is just a still frame, then the image-based error metric is adequate. In the case of view-independent solutions, the application of the metric becomes more complex because a number of "representative" views should be chosen. In practice, instead of measuring the image quality in the absolute terms, it is much easier to derive a relative metric which predicts the perceived differences between a pair of images [21]. (It is well-known that a common mean-squared error metric usually fails in such a task [5, 24, 21, 8].) A single numeric value might be adequate for some applications such as monitoring of the perceptually measured simulation progress; however, for more specific guiding of the lighting computations, a local metric operating at the pixel level is required. In this work, we investigate the application of such local metrics in the framework of the **H**ierarchical (link-less), cluster-based **P**rogressive **R**adiosity (HPR) [18] and **M**onte **C**arlo **P**ath **T**racing (MCPT) [22, 12] algorithms.

2 Modeling Human Visual System

So far, a comprehensive and complete model of Human Visual System (HVS) has not been developed. While the higher order visual processing remains mostly unexplored, the early stages in the visual pathway (beginning with the retina, ending with the visual cortex [V1], and including various intervening nuclei) are relatively well understood [6]. Since other cortical visual areas of the brain appear to receive visual information mostly via V1 neurons, an important issue becomes the representation of an image by cells in the V1. In this section, we outline such an internal representation based on spatial frequency and orientation channels, which provide the core of the most recent HVS models that attempt to describe spatial vision. Also, we discuss the existing HVS models specialized in predicting the perceived differences between images.

2.1 Cortical image representations

Though we ultimately require only a model of how visual information is processed, it is instructive to examine results of studies on organization of the retina and visual cortex, and these studies have exerted a great influence upon attempt to model human spatial vision. The anatomical, physiological and psychophysical findings suggest that receptive fields of the V1 simple cells (about 50% of all V1 cells) can be well modeled by a set of differently oriented and narrowly tuned spatial filters (channels) [16, 26, 28]. A model of these channels can be used to predict the perceptual response to visual images, and most of the HVS computational models assume that channels form the basis of internal image representation. Indeed, the channel model explains well many visual phenomena such as: the overall behavioral Contrast Sensitivity Function (CSF) (visual system sensitivity is a function of the spatial frequency and orientation content of the stimulus pattern), spatial masking (detectability of a particular pattern is reduced by the presence of a second pattern of similar frequency content), and sub-threshold summation (adding two patterns of sub-threshold contrast together can improve detectability within a common channel). See [7] for an excellent discussion of these phenomena and further references.

The characteristics of these channels subserving spatial vision are difficult to establish precisely because different psychophysical measurement methods often lead to significantly different results, and the measurement error within such experiments can be quite large. Indeed, significant day-to-day variation of measurements for a single subject have been observed. (See [6], page 205 for channel bandwidth estimates derived using various psychophysical and electrophysiological measurement techniques.) In practice, some HVS models place more emphasis upon the resemblance of the channels to the action of the visual cortex, while others place more attention on computational efficiency. In the first group, Gabor functions [16] are widely used (see [23] for further references), because of their close resemblance to receptive fields of the primate V1 cells. Also, a good match with physiological measurements are filters shaped according to DOG (differences of Gaussians) [28, 14, 7] and DOOG (differences of offset differences of Gaussians) [15] functions. Watson [26, 5] proposed to build a model, termed the Cortex transform, from filters whose shape roughly approximate 2D Gabor functions, and which can be organized into a computationally-efficient multiscale pyramid. In addition, wavelet transforms, though not targeted to approximate receptive fields of the V1 cells, have been used because of their computation efficiency and compact storage [17, 31, 8].

In practice, it seems that the choice of a particular filter function is not critical as long as it roughly approximates the point-spread function of the V1 cells [15], and the frequency and orientation bandwidths are not too broad. For all the above channel models, usually 4–6 bands are used, each with full frequency bandwidth at half height falling within the 1.0–1.6 octave range, and each exhibiting one of 3–6 different orientations. However, Teo and Heeger [24] reported a problem when trying to fit their HVS model to psychophysical data. A simple model based on only three orientation channels was not adequate, but six orientations gave a close fit. Of course, there are other factors that might be considered. For example, the HVS channels are not truly independent, but rather have been found to be mutually inhibitory [6], and the visual response to a complex stimulus (e.g., a realistic synthetic image) often involves significant inhibition between channels. While most of the existing HVS models ignore this problem, some heuristics extending the HVS models are available [15, 24]. What features are included in a given model must depend upon the application.

2.2 Visible differences predictors

Our application of the HVS model concerns how to predict whether a visible difference will be observed between two images. Therefore, we were most interested in the HVS models developed for similar tasks [30, 14, 17, 5, 24, 4, 27, 7, 8, 23], which arise from studying lossy image compression, evaluating dithering algorithms, designing CRT and flat-panel displays, and generating computer graphics. Let us now describe briefly the Visual Differences Predictor (VDP) developed by Daly [5] as a representative example, which was selected by us for our experiments on global illumination algorithms. Figure 1 contains a block diagram for the VDP. The two images to be compared (target and mask) undergo an identical initial processing. At first, the original pixel intensities are compressed by the amplitude non-linearity based on the local luminance adaptation and simulating the Weber's law-like behavior. Then processing of CSF is performed that takes into account the global state of luminance adaptation, orientation, image size and eccentricity from the fovea region. The resulting image is decomposed into the spatial frequency and orientation channels using the Cortex transform, and elevation of the detection threshold based on masking is calculated. This detection threshold is then used to normalize the contrast differences between

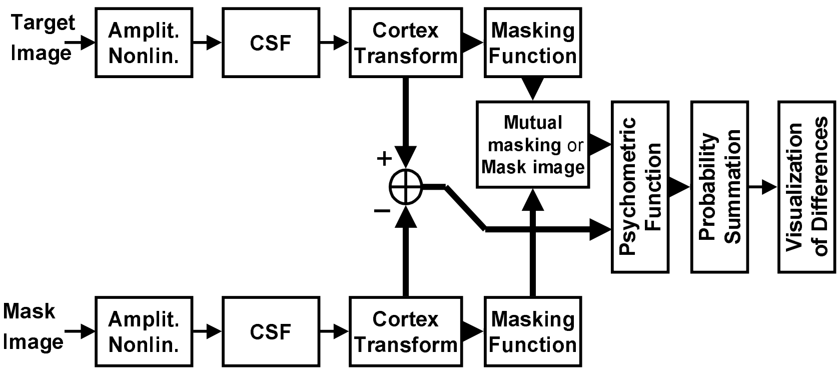


Fig. 1. Block diagram of the Visible Differences Predictor (heavy arrows indicate parallel processing of the spatial frequency and orientation channels)

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

The main advantage of the VDP (and the main reason why we choose it) is prediction of local differences between images (on the pixel level), while most of methods, including that recently developed in [8, 7], did not show such functionality, and provided a single scalar value as a measure of the differences. This locality of predictions is instrumental in developing perception-driven refinements in global illumination calculations. The Daly model also takes into account the visual characteristics that we think are extremely important in our application: a Weber's law-like amplitude compression, advanced CSF model, and masking (the latter one is ignored by many HVS models [14, 17, 8, 23]). The Cortex transform is a pyramid-style, invertible, and computationally efficient image representation, and the shape of basic filters resembles the shape of Gabor functions (which appear to be the most accurate model of receptive fields of the V1 simple cells).

Disadvantages of the Daly model have been identified, particularly from the standpoint of integrity. As pointed out by Taylor *et al.* [23], a variety of psychophysical data that is used in the Daly model was derived from the results of various unrelated experiments, usually conducted using completely different tasks. Taylor *et al.* executed psychophysical experiments that directly determined the parameters of their model. They measured contrast sensitivity as a function of frequency, luminance, and orientation of Gabor patches, which are the same functions as those used by them in their image pyramid decomposition. Also, they pointed out that it is more appropriate to use contrast discrimination thresholds instead of the more commonly used detection thresholds for predicting visible differences between images. However, they had not yet included masking into their model, which was left for future work.

The HVS model proposed by Wilson and Gelb [29] also has many attractive properties. It has good integrity since it was derived from the averaged masking data for three subjects, and shows a good fit to some adaptation and subthreshold summation experiments. Wilson and Gelb adjust the absolute sensitivity of each channel to fit threshold-sensitivity data for each of the different types of patterns and viewing conditions. However, in our application such procedure does not seem to be practical because we deal with diversified and complex stimuli in rendered images, and the results obtained in experiments employing simple and isolated stimuli are typically hard to generalize [6]. The quest for a more automatic procedure motivated Ferwerda *et al.* to extend Wilson's HVS model relying on the the results from disparate studies (perhaps at expense of the model integrity). Ferwerda *et al.* incorporated a general CSF [17] in the selection of specific channel sensitivities in order to take into account luminance adaptation under the given viewing conditions for a particular image.

The original Daly model did not support color, although its extension seems straightforward, based upon solutions proposed in [14, 17, 2, 7]. In our research, we use the VDP on an achromatic channel only. We did not want to add more heuristics to this model, which would be necessary for color processing because of a rather small volume of experimental data on the chromatic mechanisms (e.g., the luminance dependence of the chromatic CSF [17]). Also, in global illumination calculations many important effects such as shadow quality, can be relatively well captured by the achromatic mechanism, the most sensitive mechanism subserving human spatial vision.

The Daly model clearly is not perfect for reasons discussed above. Also, it ignores local luminance adaptation of the eye (only global contrast is computed), mutual channel inhibition, and suphrathreshold vision (for the most part). Still, it seems to be one of the best existing choices for our applications. One of the goals of this work was to check suitability of Daly's model to detect differences between images which arise in typical global illumination and rendering computations. What follows is a description of our experiments designed to test the validity of Daly's model in these applications.

3 VDP validation experiments

Before attempting to use Daly's VDP model in the actual global illumination computations, we chose to execute pilot studies to validate its performance in comparing realistic synthetic images. In all the experiments reported in this section, we analyzed the predicted visible differences in shadow quality for two subsequent stages of the adaptive mesh subdivision algorithm. The shadow reconstruction was a part of the view-independent illumination map computations performed using our progressive radiosity method, HPR, which begins with direct lighting computations and then proceeds with indirect lighting computations [18].

Perceived shadow quality vs. masking by texture. The goal of the first experiment was to investigate shadow masking by texture. Figure 2a and b show the reconstructed shadows cast on the floor at two different stages of mesh computations (Figure 4d in Appendix/Color Section shows the complete scene from which this region of interest was extracted). We experimented with various texture patterns and texture scales. We used images of resolution 512×512, and we assumed that the CRT was observed from the distance of 0.5 meter. Figure 2c shows the output image of the VDP when images in Figure 2a and b were used as the target and mask. Via alpha blending, color is added to each pixel in the original grey-scale image in order to indicate its difference-detection probability value (refer to [1] for the full resolution color images); red pixels indicate probability values greater than 0.75 (standard threshold value for discrimination tasks [28]), while shades of yellow correspond to probabilities below this threshold. Figure 3a summarizes the numeric VDP responses (percentage of pixels for which the probability of the difference detection is over 0.75) as a function of texture scale (the generic texture pattern is the same as that in the top row of Figure 2). The greatest shadow masking was predicted for similar scales of major texture and shadow elements (scale may be quantified in terms of spatial frequency content). Figure 3b shows VDP response as a function of texture orientation. Here, the greatest masking was predicted for texture elements aligned with major shadow elements. On the Web page [1], we provide more extended documentation of these tests for three textures. It is also possible to download the representative original images for which the experiment was performed. As it can be seen in [1], the VDP predicted less masking for textures containing more randomly oriented patterns (without dominant elements that might be aligned with shadows).

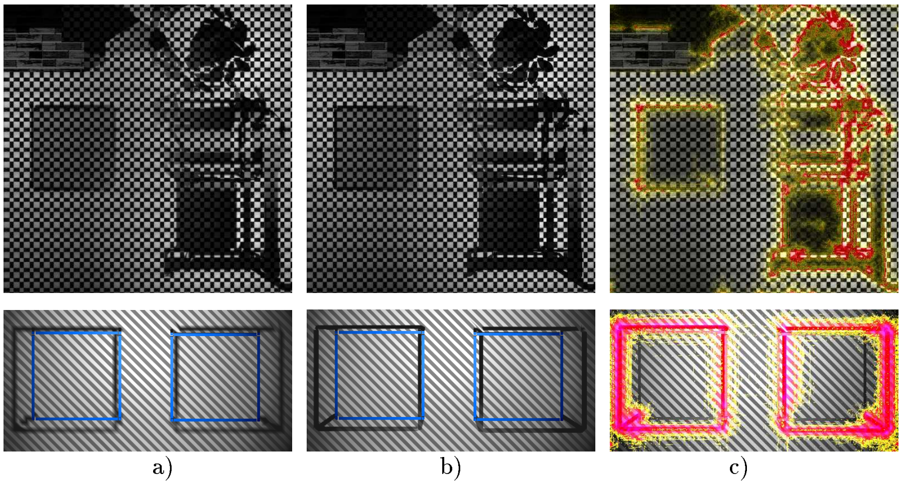


Fig. 2. Perceived shadows artifacts: a) mask, and b) target images, c) the VDP output. Refer to [1] for the full resolution color images.

Perceived shadows quality vs. observation distance. When a typical desktop CRT of resolution 1280 × 1024 is observed from the distance of 0.5 meter, the maximum number of cycles per visual degree (cpd) that can be properly reconstructed is about 15, according to the Nyquist law. This means that the distance of about 1.5 meters is needed to observe the effects of the high frequency cut-off of the CSF, which falls in the range 40–50 cpd under typical CRT viewing conditions. Figure 3c shows a graph with the VDP responses as a function of distance (see [1] for visualization of the VDP output).

Perceived shadows quality vs. solution convergence. We tested performance of the VDP in predicting visible differences between shadows for a range of illuminations resulting from progressive stages of indirect lighting computation. Figure 3d shows a graph of the VDP responses for stages ranging from the direct lighting solution (at time zero), through subsequent stages of the progressive radiosity solution. The differences in shadow visibility decrease as the solution converges, with a dramatic jump in contrast masking as the first iteration of indirect lighting is completed (see [1] for more details). Figure 3d shows for comparison the corresponding RMS error (dashed line), which changes only slightly, and poorly predicts the perceived differences. This experiment quantitatively demonstrates that important savings in terms of the number of mesh elements can be achieved if the meshing algorithm is based upon an approximated global illumination solution, even if the approximation is quite rough. This is in contrast with usual practice, in which meshing is often done based only upon the direct lighting distribution. One notable exception is a meshing technique proposed by Gibson and Hubbold [9], that uses an "ambient correction term" to compensate for the lack of knowledge of global illumination during the initial stages of the progressive radiosity solution.

4 VDP integrity

We chose to examine the Daly model integrity in terms of how critical given components were to maintaining a useful output. By replacing various model components with functionally similar components derived from results on different tasks, we were able to show how robust this model is. We experimented with three types of CSF used in the following HVS models: [5], [17, 7], and [8]. The response of the VDP was very similar in the former two cases, while for the latter one discrepancies were more significant. A possible reason for such discrepancies is that the CSF used in [8] does not take into account luminance adaptation for our test, which could differ from the conditions under which the CSF was originally measured.

We were also curious how the channel decomposition method influenced the VDP responses. We compared the Cortex transform [5], having 6 spatial frequency and 6 orientation channels, with the DOG-like pyramid construction of Burt [3], having 6 spatial frequency channels, and extended to include 4 orientation channels. While the quantitative results are different, the distribution of probabilities of detection differences between images corresponds quite well. The quantitative differences can be reduced by an appropriate scaling of the VDP responses.

Daly's original VDP model used an average image mean to compute the global contrast for every channel of the Cortex transform. We experimented with the local contrast using a low-pass filter on the input image to provide an estimate of luminance adaptation for every pixel. This made the VDP more sensitive to differences in dark image regions, and we found that in many cases the VDP responses matched better our subjective impressions (i.e., predicted

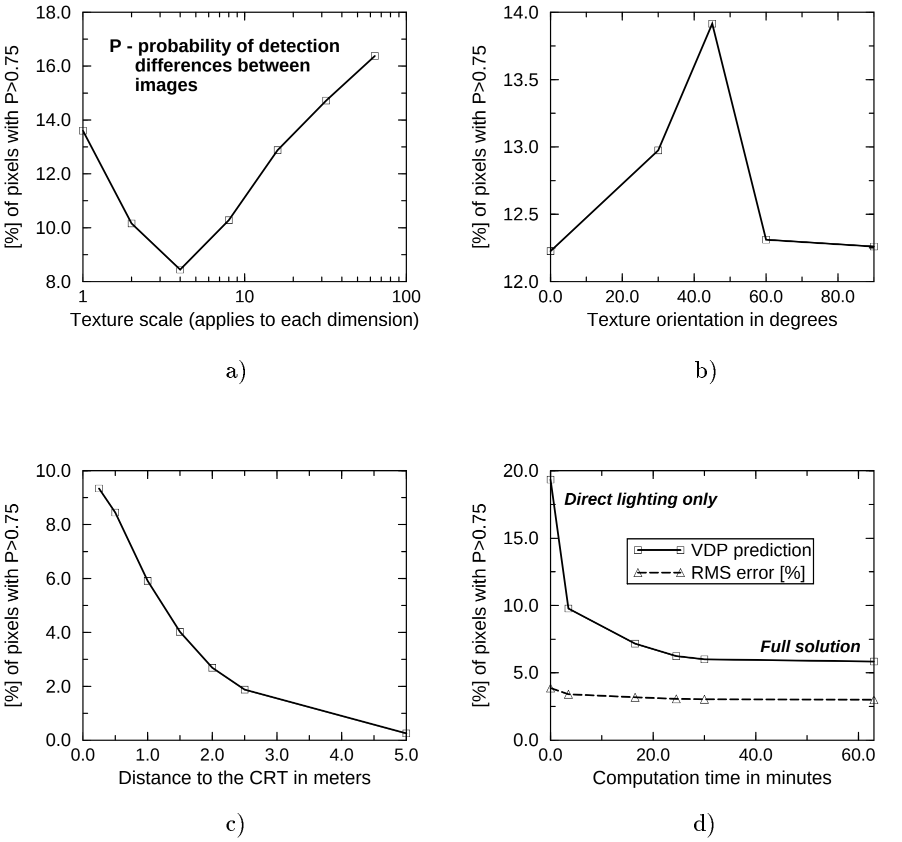


Fig. 3. Perceived shadow artifacts: masking by textures of various a) scales, and b) orientations; as a function of c) distance to the CRT (optical blur is ignored), and d) convergence of the indirect lighting solution (Note that the symbols plotted on the y axis, at time zero, indicate the probability for the direct lighting solution only, providing a reference for the subsequent stages of indirect lighting).

image differences corresponded well with actual perceived differences). In [1] we provide a variety of selected experimental results that support the conclusions reported in this section.

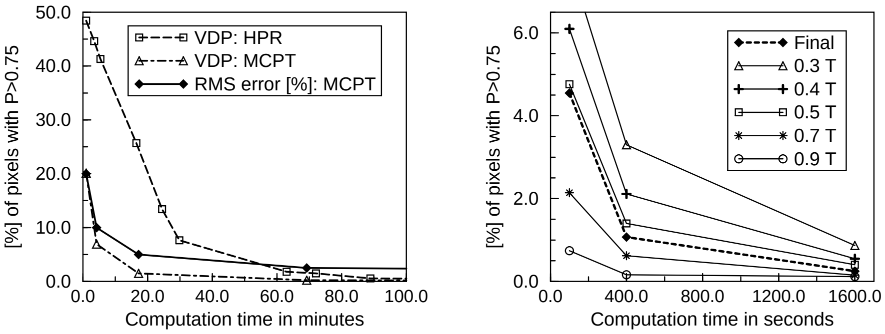
5 VDP applications

We applied the VDP to more practical tasks arising in global illumination calculations. In this section, we present results of comparison of the "perceptual" convergence of two progressive indirect lighting simulation techniques. Then we discuss perceptually-driven stopping conditions for such progressive computations. Finally, we experiment with simplification of the adaptive mesh subdivision by direct embedding of the VDP into decision process, which controls the mesh refinement.

Comparing progression of indirect lighting solution. A practical rendering system must be fast, and provide progressively updated images to show the current status of computations. A meaningful metric of the rendering progression can be expressed as perceived differences between the intermediate and final images as a function of time. In developing new or experimental illumination algorithms, we can use this metric to compare their performance to that of other solutions. Traditionally, such comparisons required the use of human observers examining the differences between the images. Using the VDP, quantitative measures of such differences are automatically generated, and also the image regions are identified in which such differences will be seen by the human observer at some high probability.

We applied the VDP to compare progressivity of indirect lighting solution for two different algorithms: the hierarchical (link-less) cluster-based progressive radiosity (HPR) [18] and Monte Carlo Path Tracing (MCPT) [12]. Both methods are based on the same ray-tracing kernel which is used to compute form factors (visibility computations) and tracing photons. We compared the ray traced images that take into account indirect lighting computed using the two methods, because we wanted to capture the influence of indirect lighting on the final appearance of specular and view-dependent effects. Figure 4a (see Appendix/Color Section) shows the intermediate images with indirect lighting computed at 1 and 30 minutes, and at 1 and 4 minutes, using the HPR and MCPT techniques (CPU R10000; the rendered scene is built of about 90,000 meshed triangles). Images in the second column (Figure 4b) show the absolute differences (normalized by the image mean) between the intermediate images (Figure 4a) and the final images (Figure 4d and e). The last column of images (Figure 4c) shows the corresponding responses of the VDP, which provide a more reliable picture of the perceived differences. This becomes especially evident for the MCPT images, which locally exhibit many fluctuations of indirect lighting that are usually well masked by textures. All such fluctuations are immediately reported by the absolute differences metrics, while the VDP responds more selectively by taking into account their contrast, spatial frequencies and masking textures. The graph in Figure 5a summarizes results of our comparisons, and shows that in our implementation the HPR technique converges much more slowly than the MCPT. (We did not use the ambient light approximation or overshooting techniques, because we are interested in physically sound intermediate results.)

Stopping conditions for lighting computations. When global illumination calculations are performed, the problem of their stopping conditions naturally presents itself. In an ideal case, computations should be stopped immediately when the image quality becomes indistinguishable from that of the fully converged solution for the human observer. (This assumption is valid if the only goal of computations is a computer image, not numerical values of illumination which can be important in some engineering applications.) Typically, the fully converged solution is not known. In practice, to stop computations, some estimate of the simulation error in terms of energy is provided by a lighting simulation algorithm, and compared against a threshold value imposed by the user. However, energy-based error measures are not reliable in predicting possible differences in appearance of the intermediate and final images. For example, the MCPT has a stochastically derived estimate of the RMS energy error [12], which

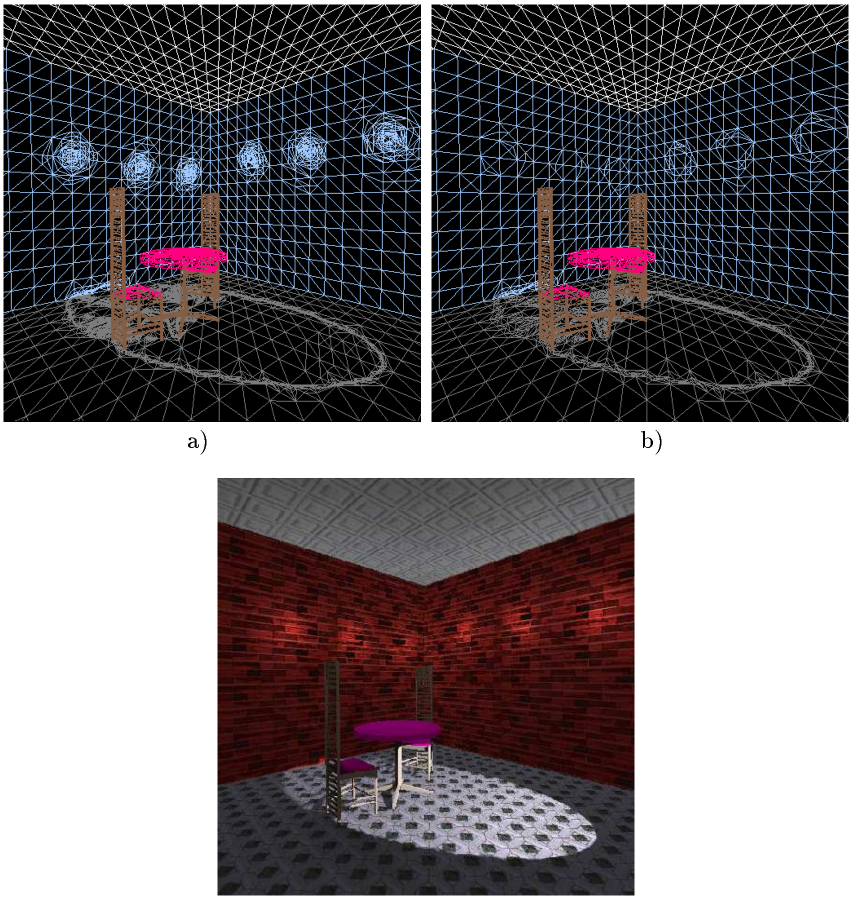


a) b)

Fig. 5. a) The VDP predicted "perceptual" convergence of indirect lighting solution for HPR and MCPT techniques. b) The VDP predicted differences between I_C and I_T, and I_T and $I_{\alpha T}$ images.

is shown in Figure 5a (solid line) for a scene in Figure 4. The actual RMS error measured between the intermediate and final images is usually significantly different because of the compressive power of the Tone Mapping Operator (TMO) [25] used to convert luminance to displayable RGB. Also, the RMS measure is global for the whole scene, and the actual error may be locally much higher. This observation led us to use images directly, and compute the differences between intermediate images using the VDP, which captures such local errors much better. If the VDP does not report significant differences between intermediate images, we assume that computations can be stopped. A question arises: *which images should be compared against the current image to get robust stopping conditions?*

Let us assume that the current image I_T is obtained after the computation time T, and let us denote by $VDP(I_T, I_{\alpha T})$ the VDP response for a pair of images I_T and $I_{\alpha T}$ where $0 < \alpha < 1$. We should find an α to get a reasonable match between $VDP(I_T, I_{\alpha T})$ and $VDP(I_C, I_T)$, where I_C is an image for the fully converged solution. We tried to address this algorithm-dependent question experimentally for the MCPT technique. Figure 5b shows the numerical values of $VDP(I_C, I_T)$ and $VDP(I_T, I_{\alpha T})$ for $T = \{100, 400, 1600\}$ seconds and various α, (the same scene as in Figure 4, but built of 17,500 meshed triangles). While the numerical values of $VDP(I_T, I_{0.5T})$ provide the upper bound for $VDP(I_C, I_T)$ over all investigated T, it is even more important that the image regions with the perceivable differences are similar in both cases (refer to [1] for color images with $VDP(I_C, I_T)$ and $VDP(I_T, I_{0.5T})$). This means that for certain regions of $I_{0.5T}$ and I_T the variance of the luminance estimate is very small (below the perceived level), and it is likely that it will be so for I_C. For other regions such variance is high, and it is likely that luminance estimates for $I_{0.5T}$ and I_T which fluctuate around the converged values for I_C will be different, and can be captured by the VDP. Thus, the choice of α is a trade-off. The α should be small enough to capture such perceivable fluctuations. However, it cannot be too small because $I_{\alpha T}$ may exhibit high variance in the regions in which the solution for I_T converged to that of I_C, with luminance differences below the noticeable level. In our experiments with stopping conditions for the MCPT technique for various scenes we found that $\alpha = 0.5$ (50% of photons are the same for I_T and $I_{0.5T}$) is such a reasonable trade-off.



a) b) c)

Fig. 6. Mesh simplification: a) original technique, b) the VDP-enhanced solution, c) the resulting shading quality.

Adaptive mesh subdivision. The goal of the adaptive meshing is to reconstruct lighting function without visible artifacts using the minimal number of mesh elements. The reported perception-based meshing solutions focus on application of Tone Mapping Operator [25] and perceptually uniform color spaces to compute perceivable differences guiding the mesh subdivision [19, 9, 11]. However, masking lighting patterns by textures, which can lead to important savings in the number of mesh elements, seems to be mostly unexplored in the existing solutions. We applied the VDP to support decisions of mesh splitting by expanding our earlier iterative mesh subdivision technique [19]. Our current implementation consists of three phases, which are repeated for every iteration: (1) choosing candidate edges for subdivision using the traditional approach [19], (2) verification of chosen candidates using the VDP, (3) restoring unnecessarily split edges. In the first phase, since some edges can become locked for subdivision as a result of previous processing, the subdivision of only active edges is decided based on differences in RGB along the edge. (We use the TMO [25] to derive the displayable RGB from luminance values, and we choose different, predefined RGB thresholds for dark and bright image regions to decide upon an edge subdivision [19].) If the differences are too small, the edge becomes locked and will be skipped in lighting calculations. At this stage, textures are ignored, and we use "representative" (averaged) color of a texture to compute RGB values. In the second phase, the mesh based on subdivided edges is built, and an image of the resulting luminance map with textures is generated. The VDP is used to compare this image with the final image from the previous iteration. The item buffer is applied to identify all triangles with visible differences (at least one pixel with $P > 0.75\%$), and whose unlocked edges are marked as candidates for subdivision. In the third phase, all active edges are re-scanned: the marked edges are subdivided, and the remaining edges are locked for future subdivisions. Finally, the perceptually adjusted mesh is built, and the final image for this iteration is generated for the VDP processing in the next iteration. Figure 6 illustrates our preliminary results. By incorporating the VDP into the mesh-splitting decision process, usually we are able to reduce the number mesh elements. Figures 6a and b show the subdivided mesh using our original and the VDP-enhanced techniques. Figures 6c shows the resulting shading which appears to be the same in both cases. The VDP-based meshing works well for smoothly changing lighting, but in the proximity of lighting discontinuities mesh reduction is rather poor due to excessive spread of predicted differences to pixels in the discontinuities neighborhood. (At high contrast, phase shifts in the channel responses can occur because of slight dislocation of the mesh-reconstructed shadow boundaries, and much wider spatial extent than the boundary dislocations themselves are required to compensate for the resulting differences in the channel responses.)

6 Conclusions

In this work, we investigated the suitability and integrity of the Visible Difference Predictor developed by Daly [5] as applied to some basic global illumination tasks. In our validation experiments we found that VDP predicts well the detectability of various target patterns (lighting solution elements such as shadows) imposed on background patterns (textures) of various scales and orientations. The integrity of the model was shown via comparisons of its predictions when the model incorporated a variety of contrast definitions, spatial and orientation channel decomposition methods, and CSFs derived from different psychophysical experiments. Though it is difficult to generalize from our results to a wide range of input images and potential VDP applications, prediction was quite robust across the tasks we examined and variations in the configuration of VDP modules. The application of the VDP to monitor "perceptual" convergence of global illumination solutions was shown to provide a meaningful metric to compare their performance and decide upon computation stopping conditions. Also, we exploited shadow masking by texture and contrast to simplify the final mesh resulting from the mesh adaptive subdivision computations.

In future work, we plan to perform more systematic validation and calibration of the VDP in human psychophysical experiments designed specifically for our task. We also wish to incorporate the modeling of suphrathreshold vision [20] and mutual channel inhibition [15, 24], which should further improve the fit of VDP responses to the results of psychophysical experiments.

7 Acknowledgments

The author would like to thank Takehiro Tawara for help in organizing the experimental material into Web pages [1]. Special thanks to Bill Martens for stimulating discussions, and for reviewing the manuscript.

References

1. `http://www.u-aizu.ac.jp/labs/csel/vdp/` - the Web page accompanying this paper.
2. Mark R. Bolin and Gary W. Meyer. A frequency based ray tracer. In *SIGGRAPH 95 Conference Proceedings*, pages 409–418, 1995.
3. P.J. Burt. Fast filter transforms for image processing. *Computer Vision, Graphics and Image Processing*, 21:368–382, 1983.
4. S. Comes, O. Bruyndonckx, and B. Macq. Image quality criterion based on the cancellation of the masked noise. In *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, pages 2635–2638, 1995.
5. S. Daly. The Visible Differences Predictor: An algorithm for the assessment of image fidelity. In A.B. Watson, editor, *Digital Image and Human Vision*, pages 179–206. Cambridge, MA: MIT Press, 1993.
6. R.L. De Valois and DeValois K.K. *Spatial vision.* Oxford University Press, Oxford, 1990.
7. J.A. Ferwerda, S. Pattanaik, P. Shirley, and D.P. Greenberg. A model of visual masking for computer graphics. In *SIGGRAPH 97 Conference Proceedings*, pages 143–152, 1997.
8. A. Gaddipatti, R. Machiraju, and R. Yagel. Steering image generation with wavelet based perceptual metric. *Computer Graphics Forum (Eurographics '97)*, 16(3):241–251, September 1997.
9. S. Gibson and R. J. Hubbold. Perceptually-driven radiosity. *Computer Graphics Forum*, 16(2):129–141, 1997.
10. D.P. Greenberg, K.E. Torrance, P. Shirley, J. Arvo, J.A. Ferwerda, S.N. Pattanaik, E.P.F. Lafortune, B. Walter, S.C. Foo, and B. Trumbore. A framework for realistic image synthesis. In *SIGGRAPH 97 Conference Proceedings*, pages 477–494, 1997.
11. David Hedley, Adam Worrall, and Derek Paddon. Selective culling of discontinuity lines. In *8th Eurographics Workshop on Rendering*, pages 69–80, 1997.
12. A. B. Khodulev and E. A. Kopylov. Physically accurate lighting simulation in computer graphics software. In *Graphicon'96*, pages 111–119, 1996.

13. D. Lischinski, B. Smits, and D.P. Greenberg. Bounds and error estimates for radiosity. In *SIGGRAPH 94 Conference Proceedings*, pages 67–74, 1994.
14. C. Lloyd and R.J. Beaton. Design of spatial-chromatic human vision model for evaluating full-color display systems. In *Human Vision and Electronic Imaging: Models, Methods, and Appl.*, pages 23–37. SPIE Vol. 1249, 1990.
15. J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A*, 7(5):923–932, 1990.
16. S. Marcelja. Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.*, 70:1297–1300, 1980.
17. R.A. Martin, A.J. Ahumada, and J.O. Larimer. Color matrix display simulation based upon luminance and chrominance contrast sensitivity of early vision. In *Human Vision, Visual Processing, and Digital Display III*, pages 336–342. SPIE Vol. 1666, 1992.
18. K. Myszkowski and T.L. Kunii. An efficient cluster-based hierarchical progressive radiosity algorithm. In *ICSC '95*, volume 1024 of *Lecture Notes in Computer Science*, pages 292–303. Springer-Verlag, 1995.
19. K. Myszkowski, A. Wojdala, and K. Wicynski. Non-uniform adaptive meshing for global illumination. *Machine Graphics and Vision*, 3(4):601–610, 1994.
20. E. Peli. Contrast in complex images. *J. Opt. Soc. Am. A*, 7(10):2033–2040, 1990.
21. H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust. Comparing real and synthetic images: some ideas about metrics. In *Sixth Eurographics Workshop on Rendering*, pages 82–91. Eurographics, June 1995.
22. P. Shirley, B. Wade, P. M. Hubbard, D. Zareski, B. Walter, and D. P. Greenberg. Global Illumination via Density Estimation. In *Sixth Eurographics Workshop on Rendering*, pages 219–230, 1995.
23. C.C. Taylor, Z. Pizlo, J. P. Allebach, and C.A. Bouman. Image quality assessment with a Gabor pyramid model of the human visual system. In *Human Vision and Electronic Imaging III.* SPIE Vol. 3299, 1998.
24. P.C. Teo and D.J. Heeger. Perceptual image distortion. pages 127–141. SPIE Vol. 2179, 1994.
25. J. Tumblin and H.E. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, 1993.
26. A.B. Watson. The Cortex transform: rapid computation of simulated neural images. *Comp. Vision Graphics and Image Processing*, 39:311–327, 1987.
27. S.J.P. Westen, R.L. Lagendijk, and J. Biemond. Perceptual image quality based on a multiple channel HVS model. In *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, pages 2351–2354, 1995.
28. H.R. Wilson. Psychophysical models of spatial vision and hyperacuity. In D. Regan, editor, *Spatial vision, Vol. 10, Vision and Visual Disfunction*, pages 179–206. Cambridge, MA: MIT Press, 1991.
29. H.R. Wilson and D.J. Gelb. Modified line-element theory for spatial-frequency and width discrimination. *J. Opt. Soc. Am. A*, 1(1):124–131, 1984.
30. C. Zetzsche and Hauske G. Multiple channel model for the prediction of subjective image quality. In *Human Vision, Visual Processing, and Digital Display*, pages 209–216. SPIE Vol. 1077, 1989.
31. B. Zhu, A.H. Tewfik, and O.N. Gerek. Image coding with mixed representations and visual masking. In *Proc. of IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, pages 2327–2330, 1995.